

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:54:58 ; Search time 44 Seconds
(without alignments)
3146.532 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKYYTAAGSAIVLSL.....IELRLPGEVKKNSLDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 segs, 133250620 residues

Word size : 8

Total number of hits satisfying chosen parameters: 169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

```
A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	1039	21	AA12715
2	1039	100.0	1039	23	AAU75932
3	1019	98.1	1019	21	AA12722
4	1019	98.1	1019	21	AA12751
5	1019	98.1	1019	23	AAU84021
6	814	78.3	1019	21	AA12750
7	807	77.7	807	23	AAU83997
8	807	77.7	1152	23	AAU84054
9	807	77.7	1238	23	AAU84056
10	807	77.7	1365	23	AAU84057

11	807	77.7	1378	23	AAU84053	S. pneumoniae deri
12	780	75.1	780	21	AA12744	Streptococcus pneu
13	713	68.6	1019	21	AA12748	Streptococcus pneu
14	713	68.6	1019	21	AA12752	Streptococcus pneu
15	679	65.4	679	23	AAU84092	Truncated variant
16	644	62.0	644	23	AAU83996	Truncated variant
17	632	60.8	632	23	AAU83998	Truncated variant
18	612	58.9	1019	21	AA12749	Streptococcus pneu
19	578	55.6	632	23	AAU84002	Truncated variant
20	575	55.3	632	23	AAU84003	Truncated variant
21	575	55.3	632	23	AAU84013	Truncated variant
22	575	55.3	632	23	AAU84018	Truncated variant
23	573	55.1	626	23	AAU84018	Truncated variant
24	573	55.1	632	23	AAU84014	Truncated variant
25	573	55.1	895	23	AAU84070	S. pneumoniae deri
26	573	55.1	896	23	AAU84080	S. pneumoniae deri
27	573	55.1	901	23	AAU84068	S. pneumoniae deri
28	573	55.1	901	23	AAU84072	S. pneumoniae deri
29	573	55.1	901	23	AAU84074	S. pneumoniae deri
30	573	55.1	902	23	AAU84076	S. pneumoniae deri
31	573	55.1	902	23	AAU84078	S. pneumoniae deri
32	568	54.7	568	21	AA12731	Streptococcus pneu
33	568	54.7	568	23	AAU83999	Truncated variant
34	568	54.7	913	23	AAU84059	S. pneumoniae deri
35	568	54.7	999	23	AAU84051	S. pneumoniae deri
36	568	54.7	999	23	AAU84052	S. pneumoniae deri
37	568	54.7	1057	21	AA12725	Streptococcus pneu
38	568	54.7	1058	23	AAU84097	S. pneumoniae deri
39	568	54.7	1126	23	AAU84058	S. pneumoniae deri
40	568	54.7	1139	23	AAU84055	S. pneumoniae deri
41	531	51.1	632	23	AAU84006	Truncated variant
42	528	50.8	528	21	AA12719	Streptococcus pneu
43	528	50.8	528	23	AAU84024	Truncated variant
44	509	49.0	509	21	AA12724	Streptococcus pneu
45	509	49.0	509	23	AAU84023	Truncated variant
46	489	47.1	489	21	AA12723	Streptococcus pneu
47	489	47.1	489	23	AAU84022	Truncated variant
48	489	47.1	632	23	AAU84004	Truncated variant
49	489	47.1	632	23	AAU84009	Truncated variant
50	489	47.1	632	23	AAU84012	Truncated variant
51	489	47.1	901	23	AAU84066	S. pneumoniae deri
52	489	47.1	901	23	AAU84067	S. pneumoniae deri
53	489	47.1	907	23	AAU84060	S. pneumoniae deri
54	489	47.1	907	23	AAU84061	S. pneumoniae deri
55	489	47.1	907	23	AAU84064	S. pneumoniae deri
56	489	47.1	907	23	AAU84065	S. pneumoniae deri
57	486	46.8	568	23	AAU84008	Truncated variant
58	486	46.8	632	23	AAU84005	Truncated variant
59	486	46.8	632	23	AAU84007	Truncated variant
60	486	46.8	632	23	AAU84015	Truncated variant
61	486	46.8	632	23	AAU84016	Truncated variant
62	486	46.8	632	23	AAU84017	Truncated variant
63	486	46.8	907	23	AAU84062	S. pneumoniae deri
64	486	46.8	907	23	AAU84063	S. pneumoniae deri
65	484	46.6	568	23	AAU84000	Truncated variant
66	484	46.6	620	23	AAU84020	Truncated variant
67	484	46.6	626	23	AAU84019	Truncated variant
68	484	46.6	632	23	AAU84001	Truncated variant
69	484	46.6	889	23	AAU84071	S. pneumoniae deri
70	484	46.6	890	23	AAU84081	S. pneumoniae deri
71	484	46.6	895	23	AAU84069	S. pneumoniae deri
72	484	46.6	895	23	AAU84073	S. pneumoniae deri
73	484	46.6	895	23	AAU84075	S. pneumoniae deri
74	484	46.6	896	23	AAU84077	S. pneumoniae deri
75	484	46.6	896	23	AAU84079	S. pneumoniae deri
76	479	46.1	484	21	AA12722	Streptococcus pneu
77	479	46.1	484	21	AA12751	Streptococcus pneu
78	479	46.1	485	21	AA12721	Streptococcus pneu
79	475	45.7	840	21	AA12721	Streptococcus pneu
80	475	45.7	840	23	AAU75151	Streptococcus pneu
81	447	43.0	447	19	AAW63228	Streptococcus pneu
82	447	43.0	447	23	ABP54647	S. pneumoniae Sfl0
83	378	36.4	484	21	AA12718	Streptococcus pneu

84	329	31.7	329	21	AA12732	Streptococcus pneu	157	8	0.8	320	22	AA599744	Oryza sativa perox
85	329	31.7	329	23	AAU84045	Truncated variant	158	8	0.8	326	22	AB564432	Drosophila melanog
86	240	23.1	240	21	AA12733	Streptococcus pneu	159	8	0.8	339	20	AA141387	Human secreted pro
87	240	23.1	240	21	AAU84046	Truncated variant	160	8	0.8	345	23	AB884747	DNA polymerase III
88	214	20.6	214	23	AAU84095	Truncated variant	161	8	0.8	514	21	AA32514	S. lavendulae MmCA
89	205	19.7	205	21	AA12726	Streptococcus pneu	162	8	0.8	514	22	AA25917	Human protein sequ
90	205	19.7	205	23	AAU84025	Truncated variant	163	8	0.8	543	20	AAW88403	Human adult testis
91	112	10.8	112	23	AAU84096	Truncated variant	164	8	0.8	543	21	AAV70016	Human protease and
92	94	9.0	94	23	AAU83827	S. pneumoniae anti	165	8	0.8	543	22	AA504382	Human cancer assoc
93	78	7.5	78	23	AAU83835	S. pneumoniae anti	166	8	0.8	745	20	AA127348	Group B Streptococ
94	57	5.5	57	23	AAU83833	S. pneumoniae anti	167	8	0.8	1100	22	ABB71772	Drosophila melanog
95	51	4.9	51	23	AAU83836	S. pneumoniae anti	168	8	0.8	1134	22	AAE00564	Human protein tyro
96	42	4.0	42	23	AAU83828	S. pneumoniae anti	169	8	0.8	1570	23	ABB97344	Novel human protei
97	40	3.8	40	23	AAU83834	S. pneumoniae anti							
98	39	3.8	39	23	AAU83829	S. pneumoniae anti							
99	36	3.5	36	23	AAU83830	S. pneumoniae anti							
100	35	3.4	35	23	AAU83832	S. pneumoniae anti							
101	32	3.1	324	21	AA12728	Streptococcus pneu							
102	32	3.1	334	23	AAU84027	Truncated variant							
103	32	3.1	590	21	AA12745	Truncated variant							
104	32	3.1	690	21	AAU84093	Truncated variant							
105	32	3.1	796	19	AAW55090	Streptococcus pneu							
106	32	3.1	796	23	AAU54584	S. pneumoniae SP03							
107	32	3.1	805	21	AA12764	Streptococcus pneu							
108	32	3.1	807	21	AA12765	Streptococcus pneu							
109	32	3.1	811	21	AA12760	Streptococcus pneu							
110	32	3.1	811	21	AA12761	Streptococcus pneu							
111	32	3.1	811	21	AA12762	Streptococcus pneu							
112	32	3.1	811	21	AA12763	Streptococcus pneu							
113	32	3.1	816	21	AA12756	Streptococcus pneu							
114	32	3.1	816	21	AA12757	Streptococcus pneu							
115	32	3.1	816	21	AA12758	Streptococcus pneu							
116	32	3.1	819	21	AA12740	Streptococcus pneu							
117	32	3.1	819	21	AA12754	Streptococcus pneu							
118	32	3.1	819	21	AA12754	Streptococcus pneu							
119	32	3.1	819	21	AA12754	Streptococcus pneu							
120	32	3.1	819	21	AA12754	Streptococcus pneu							
121	32	3.1	820	21	AA12755	Streptococcus pneu							
122	32	3.1	821	21	AA12727	Streptococcus pneu							
123	32	3.1	821	21	AA12766	Streptococcus pneu							
124	32	3.1	821	23	AAU84026	Truncated variant							
125	32	3.1	826	21	AAU84026	Truncated variant							
126	32	3.1	827	21	AAU81662	Streptococcus pneu							
127	32	3.1	834	21	AA12759	Streptococcus pneu							
128	32	3.1	838	21	AA12720	Streptococcus pneu							
129	32	3.1	838	21	AA12720	Streptococcus pneu							
130	32	3.1	838	23	AAU75934	Streptococcus pneu							
131	32	3.1	840	21	AA12716	Streptococcus pneu							
132	32	3.1	840	21	AAU75933	Streptococcus pneu							
133	32	3.1	840	21	AAU75933	Streptococcus pneu							
134	28	2.7	763	19	AAW55095	S. pneumoniae deri							
135	28	2.7	763	19	AAW55095	S. pneumoniae deri							
136	24	2.3	763	23	AAU83831	S. pneumoniae SP04							
137	21	2.0	466	23	AAU83831	S. pneumoniae anti							
138	19	1.8	721	20	AAU83831	S. pneumoniae anti							
139	16	1.5	381	21	AAU83831	S. pneumoniae anti							
140	16	1.5	793	20	AAU83831	S. pneumoniae anti							
141	16	1.5	822	21	AAU83831	S. pneumoniae anti							
142	16	1.5	822	22	AAU83831	S. pneumoniae anti							
143	16	1.5	822	22	AAU83831	S. pneumoniae anti							
144	16	1.5	822	22	AAU83831	S. pneumoniae anti							
145	16	1.5	822	23	AAU83831	S. pneumoniae anti							
146	16	1.5	822	23	AAU83831	S. pneumoniae anti							
147	16	1.5	822	23	AAU83831	S. pneumoniae anti							
148	16	1.5	822	23	AAU83831	S. pneumoniae anti							
149	16	1.5	822	23	AAU83831	S. pneumoniae anti							
150	16	1.5	822	23	AAU83831	S. pneumoniae anti							
151	12	1.2	163	20	AAU83831	S. pneumoniae anti							
152	12	1.2	168	21	AAU83831	S. pneumoniae anti							
153	12	1.2	877	23	AAU83831	S. pneumoniae anti							
154	8	0.8	100	21	AAU83831	S. pneumoniae anti							
155	8	0.8	100	21	AAU83831	S. pneumoniae anti							
156	8	0.8	144	22	AAU83831	S. pneumoniae anti							

ALIGNMENTS

RESULT 1

ID	AA12715	standard; Protein; 1039 AA.
XX	AA12715	
AC	AA12715	
XX	21-NOV-2000	(first entry)
DT	21-NOV-2000	(first entry)
XX	Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.	
DE	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.	
XX	Streptococcus pneumoniae.	
OS	WO200039299-A2.	
XX	06-JUL-2000.	
PD	20-DEC-1999;	99WO-CA01218.
XX	23-DEC-1998;	98US-0113800.
PR	(BIOC-)	BIOCHEM PHARMA INC.
XX	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;	
PI	WPI; 2000-452397/39.	
XX	N-PSDB; AAA65730.	
DR	Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteremia and/or pneumonia	
XX	Claim 18; Fig 2; 106pp; English.	
PS	The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3 protein antigen.	
CC	S. pneumoniae BVH-3 protein antigen.	
XX	Sequence 1039 AA;	
SQ	Query Match 100.0%; Score 1039; DB 21; Length 1039;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1039; Conservative 0; Mismatches 0; Gaps 0;	
QY	1 MKFSKKYTAAGSAVIVSLCAYALNQHRSDNNRVSYVDGSSQSKSENLTPOQVS 60	
DB	1 MKFSKKYTAAGSAVIVSLCAYALNQHRSDNNRVSYVDGSSQSKSENLTPOQVS 60	

QY 61 QREGIAEQIVIKITDQGVYVTHSHGDRHYHYNGKVPYDALFSEELMKDPNYQLKDADIVN 120
DB 61 QREGIAEQIVIKITDQGVYVTHSHGDRHYHYNGKVPYDALFSEELMKDPNYQLKDADIVN 120
QY 121 EVKGGYIIKVDGKYVYVYLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
DB 121 EVKGGYIIKVDGKYVYVYLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
QY 181 TNDGVYVFNPAIDIEDTGNAYIVPHGGHYHYTPKSDLSASELAALAAKAHLAGKNMPSQLS 240
DB 181 TNDGVYVFNPAIDIEDTGNAYIVPHGGHYHYTPKSDLSASELAALAAKAHLAGKNMPSQLS 240
QY 241 YSTASDNTQSVAKGSKSPANKSENQSLKELYDPSAQRYSESDGLVDPKAKIISR 300
DB 241 YSTASDNTQSVAKGSKSPANKSENQSLKELYDPSAQRYSESDGLVDPKAKIISR 300
QY 301 TNGVAIPHGDHYHFTIPYSKLSALEKTAARMVPISTGSTVSTNAKPNEVYSSLSLSLN 360
DB 301 TNGVAIPHGDHYHFTIPYSKLSALEKTAARMVPISTGSTVSTNAKPNEVYSSLSLSLN 360
QY 361 PSLATTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA 420
DB 361 PSLATTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA 420
QY 421 TSPSPILNPGTSHKHEEDGYCFDANRILIAEDESQFVMSHGDHNYHFKKDLTEEQIKA 480
DB 421 TSPSPILNPGTSHKHEEDGYCFDANRILIAEDESQFVMSHGDHNYHFKKDLTEEQIKA 480
QY 481 AOKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKLDKKEERAGIMQYGVKRESIVVN 540
DB 481 AOKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKLDKKEERAGIMQYGVKRESIVVN 540
QY 541 KKNALIIYPHGDHHDADPTDEHKPVGIGHSHSNYELFKPEEGVAKGKNVYTGELINV 600
DB 541 KKNALIIYPHGDHHDADPTDEHKPVGIGHSHSNYELFKPEEGVAKGKNVYTGELINV 600
QY 601 VLLKNTSTNNQFTLANGOKRVSFPPLEKKGICINMLVLIIPDGKVLKGVSKVFG 660
DB 601 VLLKNTSTNNQFTLANGOKRVSFPPLEKKGICINMLVLIIPDGKVLKGVSKVFG 660
QY 661 EGVNTANFELDQYLPQGTFFYTIASKDYPEVSDGTFTVPTSLAYKMASOTIYPPHA 720
DB 661 EGVNTANFELDQYLPQGTFFYTIASKDYPEVSDGTFTVPTSLAYKMASOTIYPPHA 720
QY 721 GDYLVNPFQFVAVPKGTDALVRVDFEFGNAYLENNYKVGELKLPKLNQGTTRTAGNK 780
DB 721 GDYLVNPFQFVAVPKGTDALVRVDFEFGNAYLENNYKVGELKLPKLNQGTTRTAGNK 780
QY 781 IPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQPKRKAQENSKLDEKVEEPTKS 840
DB 781 IPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQPKRKAQENSKLDEKVEEPTKS 840
QY 841 EKVEKEKLSFTGNTSNSTLEEVPTDQVQKVAEESYGMKLENVLFNMDGTIELYLP 900
DB 841 EKVEKEKLSFTGNTSNSTLEEVPTDQVQKVAEESYGMKLENVLFNMDGTIELYLP 900
QY 901 SGEVVKNNMADFTGEAPQNGENKPSNGKSVGTGTVENOPTENKPADSLPEAPNEKPVKP 960
DB 901 SGEVVKNNMADFTGEAPQNGENKPSNGKSVGTGTVENOPTENKPADSLPEAPNEKPVKP 960
QY 961 ENSTDNGMLNPEGNVGSDPMLDPALEAPAVDPVQKLEKFTASYGLGLDLSVIFNMDGTI 1020
DB 961 ENSTDNGMLNPEGNVGSDPMLDPALEAPAVDPVQKLEKFTASYGLGLDLSVIFNMDGTI 1020
QY 1021 ELRLPSGEVVKNNLSDFIA 1039
DB 1021 ELRLPSGEVVKNNLSDFIA 1039

RESULT 2
AAU75932
ID AAU75932 standard; Protein; 1039 AA.

XX AAU75932;
AC 08-MAY-2002 (first entry)
XX Streptococcus pneumoniae BVH-3 protein version #1.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection.
XX Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
XX W0200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001W0-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX N-PSDB: ABK15101.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia.
XX Example 1; Fig 6; 113pp; English.
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novarum or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This is the amino acid sequence of
CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
CC peptides described in the method of the invention.
XX Sequence 1039 AA;
SQ Query Match 100.0%; Score 1039; DB 23; Length 1039;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1039; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKFSKKYTAAGSAVIVSLCAYALNQHRSEKNNRVSYVDGSSQSKSENLTDPDVS 60
DB 1 MKFSKKYTAAGSAVIVSLCAYALNQHRSEKNNRVSYVDGSSQSKSENLTDPDVS 60
QY 61 QREGIAEQIVIKITDQGVYVTHSHGDRHYHYNGKVPYDALFSEELMKDPNYQLKDADIVN 120
DB 61 QREGIAEQIVIKITDQGVYVTHSHGDRHYHYNGKVPYDALFSEELMKDPNYQLKDADIVN 120
QY 121 EVKGGYIIKVDGKYVYVYLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
DB 121 EVKGGYIIKVDGKYVYVYLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180

```
QY 181 TTNDGYVFPADIIEDTGNAYIVPHGGHHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLS 240
DB 181 TTNDGYVFPADIIEDTGNAYIVPHGGHHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLS 240
QY 241 YSSTASDNNTQSVAKGTSKPKANKSENLOSLLKELYDPSAORYSESGLVDPDAKIISR 300
DB 241 YSSTASDNNTQSVAKGTSKPKANKSENLOSLLKELYDPSAORYSESGLVDPDAKIISR 300
QY 301 TPNQVAIPRGDHHYFTIPYKLSALEPKIARMVPISTGTVSTNAKPNNEVSSLSLSSN 360
DB 301 TPNQVAIPRGDHHYFTIPYKLSALEPKIARMVPISTGTVSTNAKPNNEVSSLSLSSN 360
QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIQOPTLPNNLSA 420
DB 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIQOPTLPNNLSA 420
QY 421 TSPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEEQIKA 480
DB 421 TSPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEEQIKA 480
QY 481 AQKHELVKTSUNGDLSSSHEDDYPGNKAKEMKDLKKIEEKIAGIMKOYGVKRESIVVN 540
DB 481 AQKHELVKTSUNGDLSSSHEDDYPGNKAKEMKDLKKIEEKIAGIMKOYGVKRESIVVN 540
QY 541 KEKNAIYPRGDDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
DB 541 KEKNAIYPRGDDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
QY 601 VNLKSTNNQNTLANGCKRVSFSPFPELEKLGILNMLVLIITPDGKVLKVSQKVF 660
DB 601 VNLKSTNNQNTLANGCKRVSFSPFPELEKLGILNMLVLIITPDGKVLKVSQKVF 660
QY 661 EGVGNIANFELDQYLPFGTFRYITASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPEHA 720
DB 661 EGVGNIANFELDQYLPFGTFRYITASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPEHA 720
QY 721 GDYLRVNOFOFVPGKTDALVRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTITAGNK 780
DB 721 GDYLRVNOFOFVPGKTDALVRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTITAGNK 780
QY 781 IPTVFMANAYLDNOSTYIYVEVPTLEKENOTDKPSILPQFKRNKAQENSKLDEKVEEPTS 840
DB 781 IPTVFMANAYLDNOSTYIYVEVPTLEKENOTDKPSILPQFKRNKAQENSKLDEKVEEPTS 840
QY 841 EKVEKELSTGNTSNTSLEEVPTVDVQEVAKPAESYGMKLENVLFNMDGTIELYLP 900
DB 841 EKVEKELSTGNTSNTSLEEVPTVDVQEVAKPAESYGMKLENVLFNMDGTIELYLP 900
QY 901 SGVEIKKNMADFTGEAPQNGENKPSKENGKSVSTGTVENOPTENKPADSLPEAPNEKPKVP 960
DB 901 SGVEIKKNMADFTGEAPQNGENKPSKENGKSVSTGTVENOPTENKPADSLPEAPNEKPKVP 960
QY 961 ENSTDGMLNPEGNVSDPMLDPALEAPADVPQVEKLEKFTASTYGLGLSDSVIFNMDGTI 1020
DB 961 ENSTDGMLNPEGNVSDPMLDPALEAPADVPQVEKLEKFTASTYGLGLSDSVIFNMDGTI 1020
QY 1021 ELRLPSGEVTKNLSDFIA 1039
DB 1021 ELRLPSGEVTKNLSDFIA 1039
```

RESULT 3

AAB12722

ID AAB12722 standard; Protein; 1019 AA.

XX

AC AAB12722;

XX

DT 21-NOV-2000 (first entry)

XX

DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

```
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
FN W02000039239-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
WP; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Claim 18; Fig 20; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3M protein antigen.
XX
SQ Sequence 1019 AA;
```

Query Match 98.1%; Score 1019; DB 21; Length 1019;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNQHRSEKNNRVSVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGV 80

DB 1 CAYALNQHRSEKNNRVSVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGV 60

QY 81 TSHGDHYYHNGKVPYDALFSEELMKDPNQLKADATVNEVKGYYLIKVDGYYVYVKD 140

DB 61 TSHGDHYYHNGKVPYDALFSEELMKDPNQLKADATVNEVKGYYLIKVDGYYVYVKD 120

QY 141 AAHADNVRTKDEINRQKQEHVKDKNEKNSNVAVARSQGRYTTNDGYVFNPAITDGTNA 200

DB 121 AAHADNVRTKDEINRQKQEHVKDKNEKNSNVAVARSQGRYTTNDGYVFNPAITDGTNA 180

QY 201 YIVPHGGHHYIIPKSDLSASELAALAAKHAHLAKGNMQPSQLSSTASDNNTQSVAKGTSK 260

DB 181 YIVPHGGHHYIIPKSDLSASELAALAAKHAHLAKGNMQPSQLSSTASDNNTQSVAKGTSK 240

QY 261 PANKSENLOSLLKELYDPSAORYSESGLVDPDAKIISRTPNGVAIPHGDDHFIYIPSK 320

DB 241 PANKSENLOSLLKELYDPSAORYSESGLVDPDAKIISRTPNGVAIPHGDDHFIYIPSK 300

QY 321 LSALEEKIARMVPISTGTVSTNAKPNNEVSSLSLSSNPSLTTSKELSSASDGYIFN 380

DB 301 LSALEEKIARMVPISTGTVSTNAKPNNEVSSLSLSSNPSLTTSKELSSASDGYIFN 360

QY 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHERKEED 440

DB 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHERKEED 420

QY 441 GYGEDANRIIAEDSGFVMSHGDNHNYFFKDLTEEQIKAAQKHELVKTSUNGDLSS 500

DB 421 GYGEDANRIIAEDSGFVMSHGDNHNYFFKDLTEEQIKAAQKHELVKTSUNGDLSS 480

QY 501 HEQDYPGNAKEMKDLKKIEEKIAGIMKOYGVKRESIVVANKENAIYIPHGDDHHDPI 560

Db 481 HEQDYPGNAKEMKDLKIEEKIAGIMKQYGVKRESIVVNEKKNAILYPHGDHHPADPID 540
Qy 561 EHKPVGIGHSHSNVELFKPEGVAKKKNVYTGEEELTNVNNLLKNSFNQNTFLANGQ 620
Db 541 EHKPVGIGHSHSNVELFKPEGVAKKKNVYTGEEELTNVNNLLKNSFNQNTFLANGQ 600
Qy 621 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGVKFGGVGNIANFELDQYPLPQGT 680
Db 601 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGVKFGGVGNIANFELDQYPLPQGT 660
Qy 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPGKTDAL 740
Db 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPGKTDAL 720
Qy 741 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIYVE 800
Db 721 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIYVE 780
Qy 801 VPILKEKNQTDKPSILPQFKENKQAKNSKLDKEVEEPTSEKVEKEKLSGTGNSSTL 860
Db 781 VPILKEKNQTDKPSILPQFKENKQAKNSKLDKEVEEPTSEKVEKEKLSGTGNSSTL 840
Qy 861 BEVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
Db 841 BEVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Qy 921 GENKPSNGKYSTGTIVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGSDDPM 980
Db 901 GENKPSNGKYSTGTIVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGSDDPM 960
Qy 981 LDPALAEAPADVPQVEKLEKTASYGLDSVIFNMMDGTIELRPSGEVIKKNLSDFIA 1039
Db 961 LDPALAEAPADVPQVEKLEKTASYGLDSVIFNMMDGTIELRPSGEVIKKNLSDFIA 1019

RESULT 4
AAB12751
ID AAB12751 standard; Protein; 1019 AA.
XX AC AAB12751;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia.
XX PS Disclosure; Fig 11; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 98.1%; Score 1019; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 CAVALNOHRSQENKNNRVSYVDGSSOSSSENLTPDQVSQKEGIAEQIIVKITQGGYV 80
Db 1 CAVALNOHRSQENKNNRVSYVDGSSOSSSENLTPDQVSQKEGIAEQIIVKITQGGYV 60
Qy 81 TSHGDHYHYNGVYPYDALFSEELLKMDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLDK 140
Db 61 TSHGDHYHYNGVYPYDALFSEELLKMDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLDK 120
Qy 141 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIIEDTGN 200
Db 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIIEDTGN 180
Qy 201 YIYPHGHYHYIPKSDLSASELAARAHLAGKNWQSQLSYSTASDNNQTSQVAKGSTK 260
Db 181 YIYPHGHYHYIPKSDLSASELAARAHLAGKNWQSQLSYSTASDNNQTSQVAKGSTK 240
Qy 261 PANKSENLOQLKELYDPSAQRYSQSDGLVFPDPAKIIISRTPNGVAIPHGDHYHFIYSK 320
Db 241 PANKSENLOQLKELYDPSAQRYSQSDGLVFPDPAKIIISRTPNGVAIPHGDHYHFIYSK 300
Qy 321 LSALAEKIAKRVPISTGTVSTNAKPNVSVLSGLSSNPSLTTSKELSSASDGYIFN 380
Db 301 LSALAEKIAKRVPISTGTVSTNAKPNVSVLSGLSSNPSLTTSKELSSASDGYIFN 360
Qy 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOQTLPPNNSLATPSPSLPPIPGTSHERHEED 440
Db 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOQTLPPNNSLATPSPSLPPIPGTSHERHEED 420
Qy 441 GYGFANDRIIAEDSGFVMSGHGHNHYFFFKDLTEQIKAQKHLEEVKTSNGLDLSLS 500
Db 421 GYGFANDRIIAEDSGFVMSGHGHNHYFFFKDLTEQIKAQKHLEEVKTSNGLDLSLS 480
Qy 501 HEQDYPGNAKEMKDLKIEEKIAGIMKQYGVKRESIVVNEKKNAILYPHGDHHPADPID 560
Db 481 HEQDYPGNAKEMKDLKIEEKIAGIMKQYGVKRESIVVNEKKNAILYPHGDHHPADPID 540
Qy 561 EHKPVGIGHSHSNVELFKPEGVAKKKNVYTGEEELTNVNNLLKNSFNQNTFLANGQ 620
Db 541 EHKPVGIGHSHSNVELFKPEGVAKKKNVYTGEEELTNVNNLLKNSFNQNTFLANGQ 600
Qy 621 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGVKFGGVGNIANFELDQYPLPQGT 680
Db 601 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGVKFGGVGNIANFELDQYPLPQGT 660
Qy 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPGKTDAL 740
Db 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPGKTDAL 720
Qy 741 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIYVE 800
Db 721 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIYVE 780
Qy 801 VPILKEKNQTDKPSILPQFKENKQAKNSKLDKEVEEPTSEKVEKEKLSGTGNSSTL 860
Db 781 VPILKEKNQTDKPSILPQFKENKQAKNSKLDKEVEEPTSEKVEKEKLSGTGNSSTL 840
Qy 861 BEVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
Db 841 BEVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900

QY 921 GENKPSSENGKSTGTVENOPTENKPADSLPAPNEKPKVPKPNSTNDGMLNPEGNGVSDPM 980
DB 901 GENKPSSENGKSTGTVENOPTENKPADSLPAPNEKPKVPKPNSTNDGMLNPEGNGVSDPM 960
QY 981 LDPALPEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1039
DB 961 LDPALPEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1019

RESULT 5
AAU84021
ID AAU84021 standard; Peptide; 1019 AA.
XX
AC AAU84021;
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of *S. pneumoniae* BVH-3, BVH-3M.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
BN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
DR
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
PS Example 1; Page -; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (i) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (ii) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (iii) encoding (i) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for *S. pneumoniae* infection. (iii) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1019 AA;

Query Match 98.1%; Score 1019; DB 23; Length 1019;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNQHRSQENKDNRRVSYVDGSSSQSKSENLTPDQVSQKSGIOAEQIVIKITDQGVY 80
DB 1 CAYALNQHRSQENKDNRRVSYVDGSSSQSKSENLTPDQVSQKSGIOAEQIVIKITDQGVY 60
QY 81 TSHGCDHYHYNGVVPYDALFSEELLKMDPNYQKLDADIYNEVKGGYIIKVDGKYVYLKD 140
DB 61 TSHGCDHYHYNGVVPYDALFSEELLKMDPNYQKLDADIYNEVKGGYIIKVDGKYVYLKD 120
QY 141 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFPADIETGNA 200
DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFPADIETGNA 180
QY 201 YIVPHGGHYHYIPKSDLSASELAALAAKLAGKNMQPSQSYSTASDNNNTQSAKSTSK 260
DB 181 YIVPHGGHYHYIPKSDLSASELAALAAKLAGKNMQPSQSYSTASDNNNTQSAKSTSK 240
QY 261 PANKSENLOSLLKELYDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPHGDHYHFIPYSK 320
DB 241 PANKSENLOSLLKELYDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
QY 321 LSALKEKIARMPISGTGTVSTNAKPNNEVSSLSGLSSNPSSLTTSKELSSASDGYIFN 380
DB 301 LSALKEKIARMPISGTGTVSTNAKPNNEVSSLSGLSSNPSSLTTSKELSSASDGYIFN 360
QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 440
DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
QY 441 GYGFDANRIIAEDSGFVMSHGDNHNYFFKKDLTEPQIKAAQKHLEEVKTSNGLDLSLS 500
DB 421 GYGFDANRIIAEDSGFVMSHGDNHNYFFKKDLTEPQIKAAQKHLEEVKTSNGLDLSLS 480
QY 501 HEQDYPGNNAKEMKDLKKIEEKTAGIMKQYGVKRESIVYNKEKNALIIYPHGDHHDADPID 560
DB 481 HEQDYPGNNAKEMKDLKKIEEKTAGIMKQYGVKRESIVYNKEKNALIIYPHGDHHDADPID 540
QY 561 EHKPVGIGHSNHYELFKPEEGVAKKEGKNVYTGSELTNVVNLKKNSTFNNQNTLANGQ 620
DB 541 EHKPVGIGHSNHYELFKPEEGVAKKEGKNVYTGSELTNVVNLKKNSTFNNQNTLANGQ 600
QY 621 KRVSFSPPELEKKLGINMLVLTIPDGKYLEKSVKGVEGVCNLANFELDQYLPQGT 680
DB 601 KRVSFSPPELEKKLGINMLVLTIPDGKYLEKSVKGVEGVCNLANFELDQYLPQGT 660
QY 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDAL 740
DB 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDAL 720
QY 741 VRVDFPHGNAYLENNYKVGEEKLPIPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE 800
DB 721 VRVDFPHGNAYLENNYKVGEEKLPIPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE 780
QY 801 VPILEKENQTDKPSILPQPKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLS 860
DB 781 VPILEKENQTDKPSILPQPKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLS 840
QY 861 EEPVTPDQVEKVAKEAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNADFTGEAPQGN 920
DB 841 EEPVTPDQVEKVAKEAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNADFTGEAPQGN 900
QY 921 GENKPSSENGKSTGTVENOPTENKPADSLPAPNEKPKVPKPNSTNDGMLNPEGNGVSDPM 980
DB 901 GENKPSSENGKSTGTVENOPTENKPADSLPAPNEKPKVPKPNSTNDGMLNPEGNGVSDPM 960
QY 981 LDPALPEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1039
DB 961 LDPALPEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1019

RESULT 6
AAB12750

ID AAB12750 standard; Protein; 1019 AA.
 XX AAB12750;
 AC
 XX
 DT 21-NOV-2000 (first entry)
 DE
 DE Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI
 XX WPI: 2000-452397/39.
 DR
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 PT
 XX Disclosure; Fig 11; 106pp; English.
 PS
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SQ Sequence 1019 AA;

Query Match 78.3%; Score 814; DB 21; Length 1019;
 Best Local Similarity 99.8%; Pred. NO. 0;
 Matches 1014; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 21 CAVNALGHRQENKDNRRVSYDGSQSSQSENLTDPQVSKREGIAEQIVIKITDQGV 80
 DB 1 CAVNALGHRQENKDNRRVSYDGSQSSQSENLTDPQVSKREGIAEQIVIKITDQGV 60
 QY 81 TSHGDHYHYNGKVPYDALFSELLMKDPYQLKADIVNEVKGYYIIKVDGYYVYLKD 140
 DB 61 TSHGDHYHYNGKVPYDALFSELLMKDPYQLKADIVNEVKGYYIIKVDGYYVYLKD 120
 QY 141 AAHADNVRTKDEINRQKQEHVKDNEKNSVAVARSQGYTTNDGYVFNPAIIEDTGN 200
 DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSVAVARSQGYTTNDGYVFNPAIIEDTGN 190
 QY 201 YIVPHGGHYHYIPKSDLSASELAALAGKNNQPSQLSYSTASDNNTQSVAKGSTK 260
 DB 181 YIVPHGGHYHYIPKSDLSASELAALAGKNNQPSQLSYSTASDNNTQSVAKGSTK 240
 QY 261 PANKSENLOSLLKELYDSPAQRYSSEGLVFPDPAKIIISRTPNGVAIPHGDDHYHFIPYK 320
 DB 241 PANKSENLOSLLKELYDSPAQRYSSEGLVFPDPAKIIISRTPNGVAIPHGDDHYHFIPYK 300
 QY 321 LSALEEKIARMPISGTGVSIVSTNAKNEVYVSSLSGSSNPSSLTTSKELSSASDGYFN 380
 DB 301 LSALEEKIARMPISGTGVSIVSTNAKNEVYVSSLSGSSNPSSLTTSKELSSASDGYFN 360

QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGQFTLPNNSLATPSPSLPINPOTSHKHEED 440
 DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGQFTLPNNSLATPSPSLPINPOTSHKHEED 420
 QY 441 GYGFDANRIIAEDSGFVMSHGDHNYFFKDKLTESQTKAAQKHLVEEYKTSNGLDLS 500
 DB 421 GYGFDANRIIAEDSGFVMSHGDHNYFFKDKLTESQTKAAQKHLVEEYKTSNGLDLS 480
 QY 501 HBQDYPGNAKEMKLDKXIEEKIAGIMKOYGVKRESIVVNEKNAIYVPHGDHHAADPID 560
 DB 481 HBQDYPGNAKEMKLDKXIEEKIAGIMKOYGVKRESIVVNEKNAIYVPHGDHHAADPID 540
 QY 561 EHKPVGIGHSHSNYELFPKEGVAKKEGNKYVTGELTNVNLKNSFNQNTLANGQ 620
 DB 541 EHKPVGIGHSHSNYELFPKEGVAKKEGNKYVTGELTNVNLKNSFNQNTLANGQ 600
 QY 621 KRVSFSPPELEKLGINMLVKLITPDGKVLKSVKGVGEGVGNIAFELDQVPLPQOT 680
 DB 601 KRVSFSPPELEKLGINMLVKLITPDGKVLKSVKGVGEGVGNIAFELDQVPLPQOT 660
 QY 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIPYPFHAGDTYLRVNPQFAVPGKTDAL 740
 DB 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIPYPFHAGDTYLRVNPQFAVPGKTDAL 720
 QY 741 VRVDFEFGHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTETMANAYLDNOSTYIVE 800
 DB 721 VRVDFEFGHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTETMANAYLDNOSTYIVE 780
 QY 801 VPILKEKNTDPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 860
 DB 781 VPILKEKNTDPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840
 QY 861 EEPVTVDPVQEVKAFESYGNKLENVLFNMDGTIELYLPSEVKKKNMADFTGEAPOGN 920
 DB 841 EEPVTVDPVQEVKAFESYGNKLENVLFNMDGTIELYLPSEVKKKNMADFTGEAPOGN 900
 QY 921 GENKSENGKVGSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPGNVGSDPM 980
 DB 901 GENKSENGKVGSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPGNVGSDPM 960
 QY 981 LDPALAEAPVDPVQEKLEKFTASYGLGLDVSIFNMDGTIELRLPSGEVKKKNLSD 1036
 DB 961 LDPALAEAPVDPVQEKLEKFTASYGLGLDVSIFNMDGTIELRLPSGEVKKKNLSD 1016

RESULT 7
 AAU83997
 ID AAU83997 standard; Peptide: 807 AA.
 XX
 AC AAU83997;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW25.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 OS
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI
 XX

DR WPI: 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia

XX

PS Example 1; Page : 113pp; English.

XX

CC The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocidia or

CC Streptococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX

SQ Sequence 807 AA:

Query Match 77.7%; Score 807; DB 23; Length 807;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSVAGSKTSKPAKSENQSLKLLKELYDPSAQRYSSEGLVFP 292

DB 1 NMQPSQLSYSTASDNNQTSVAGSKTSKPAKSENQSLKLLKELYDPSAQRYSSEGLVFP 60

QY 293 DPAKIIISRTNGVAIPGHDHYHPIFYPSKLSALEEKIARMPISGTSTVSTNAKPNNEVVS 352

DB 61 DPAKIIISRTNGVAIPGHDHYHPIFYPSKLSALEEKIARMPISGTSTVSTNAKPNNEVVS 120

QY 353 SLGSLSNPSLTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 412

DB 121 SLGSLSNPSLTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 180

QY 413 TLPNNSLATPSLPIPGTSHKHEEDGYGFDPANRIIAEDSGFVMSHGDHNNHYFFKKD 472

DB 181 TLPNNSLATPSLPIPGTSHKHEEDGYGFDPANRIIAEDSGFVMSHGDHNNHYFFKKD 240

QY 473 LTESQIKAAQKHLEEVKTSINGLSLSSHQDYPPGNAKEMKMDKKIEKIAKIMQYGV 532

DB 241 LTESQIKAAQKHLEEVKTSINGLSLSSHQDYPPGNAKEMKMDKKIEKIAKIMQYGV 300

QY 533 KRESIVYVNEKKNATIIYPGHGHHADPIDEHKPVYIGHSNHYELFKPEGVAKKEGNKYV 592

DB 301 KRESIVYVNEKKNATIIYPGHGHHADPIDEHKPVYIGHSNHYELFKPEGVAKKEGNKYV 360

QY 593 TGEELTNVWLLKNSTFNQNTLANGQKRVSTFSFPELEKLGILNMLVKLIPDQGVLE 652

DB 361 TGEELTNVWLLKNSTFNQNTLANGQKRVSTFSFPELEKLGILNMLVKLIPDQGVLE 420

QY 653 KVSCKVFGEGVGNIANFELDQVLPQGTFTYTIASKDYPEVSVDCGTFVTSLAYKWSQ 712

DB 421 KVSCKVFGEGVGNIANFELDQVLPQGTFTYTIASKDYPEVSVDCGTFVTSLAYKWSQ 480

QY 713 TIFVPPFHAGDTYLRVNPQFAPVPGTADALVRFDFEFGHGNAYLNNYKVGIEKLPKPLNQG 772

DB 481 TIFVPPFHAGDTYLRVNPQFAPVPGTADALVRFDFEFGHGNAYLNNYKVGIEKLPKPLNQG 540

QY 773 TTRTAGNKIPVTEMANAYLDNQSTYIVVEVPILEKENOTDKPSILPQFKRNKAQENSKLDE 832

DB 541 TTRTAGNKIPVTEMANAYLDNQSTYIVVEVPILEKENOTDKPSILPQFKRNKAQENSKLDE 600

QY 833 KVEEPTSEKVEKEKLSSETGNSTNSTLEEYPTVDPVOEKVAKFAESYGMKLENVLFNMD 892

DB 601 KVEEPTSEKVEKEKLSSETGNSTNSTLEEYPTVDPVOEKVAKFAESYGMKLENVLFNMD 660

QY 893 GTIELYLSGGEVKKKNMADFTGEAPQNGENKPSNGKVSSTGTVENOPTENKPADSLPEA 952

DB 661 GTIELYLSGGEVKKKNMADFTGEAPQNGENKPSNGKVSSTGTVENOPTENKPADSLPEA 720

QY 953 PNEKPKVPKPNSTNDGMLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012

DB 721 PNEKPKVPKPNSTNDGMLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 780

QY 1013 IFNMDGTIELRLPSPGEVKKKLSDFIA 1039

DB 781 IFNMDGTIELRLPSPGEVKKKLSDFIA 807

RESULT 8

ASU84054

ID AU84054 standard; Peptide: 1152 AA.

AC AU84054;

XX 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, NEW27.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutein;

KW BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

PN 27-DEC-2001.

PD 19-JUN-2001; 2001WO-CA00908.

PF 20-JUN-2000; 2000US-212683P.

PR (SHIR-) SHIRE BIOCHEM INC.

PA Hamel J, Quellet C, Charland N, Martin D, Brodeur B;

PI WPI: 2002-122272/16.

XX

DR

XX

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia

XX

PS Example 1; Page : 113pp; English.

XX

CC The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocidia or

CC Streptococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX

CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1152 AA;

Query Match 77.7%; Score 807; DB 23; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NNQPSQSYSTASDNTQSVAKGSTSPANKSENLSLLKELYDSPSAQRYSESGLVLF 292
DB 346 NNQPSQSYSTASDNTQSVAKGSTSPANKSENLSLLKELYDSPSAQRYSESGLVLF 405
QY 293 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPIISGTGSTVSTNAKNEVVS 352
DB 406 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPIISGTGSTVSTNAKNEVVS 465
QY 353 SLGSSSNPSSLTTSKELSSASDGIYFNPQDIVEATATAYIVRHGDHFIYIPKSNQIQCP 412
DB 466 SLGSSSNPSSLTTSKELSSASDGIYFNPQDIVEATATAYIVRHGDHFIYIPKSNQIQCP 525
QY 413 TLPNNSLATPSPSLNPQTSHEKHEEDGYFDPANRIATAEDSGFVMSGDHNNHFFKKD 472
DB 526 TLPNNSLATPSPSLNPQTSHEKHEEDGYFDPANRIATAEDSGFVMSGDHNNHFFKKD 585
QY 473 LTEEQIKAAQKHEEVKTSNGLDSSHEQDYPGNKEMKDLKIEKIAGIMQYGV 532
DB 586 LTEEQIKAAQKHEEVKTSNGLDSSHEQDYPGNKEMKDLKIEKIAGIMQYGV 645
QY 533 KRESIVVNEKNALIIYPHGDHHAIDIEHKPVGTGCHSHSNYELKPEGVAKKEGKVY 592
DB 646 KRESIVVNEKNALIIYPHGDHHAIDIEHKPVGTGCHSHSNYELKPEGVAKKEGKVY 705
QY 593 TGEELTNVNLKNSFNQNTFLANGQKRVSFSPPELEKKGINMLVKLITPDGKYLE 652
DB 706 TGEELTNVNLKNSFNQNTFLANGQKRVSFSPPELEKKGINMLVKLITPDGKYLE 765
QY 653 KVSQKVFGEVGNIANFELDQYLPQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
DB 766 KVSQKVFGEVGNIANFELDQYLPQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 825
QY 713 TIFYPPHAGDTVLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGSEIKLPIPKLNOG 772
DB 826 TIFYPPHAGDTVLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGSEIKLPIPKLNOG 885
QY 773 TTRTAGNKIPVTFMANAYLDNOSTIYVEVPILEKENQTDKPSILQFQRNKAQENSKLDE 832
DB 886 TTRTAGNKIPVTFMANAYLDNOSTIYVEVPILEKENQTDKPSILQFQRNKAQENSKLDE 945
QY 833 KYEPTKSEKVEKLSSETNGSTNSTLEVPDVPQEKVAKFAESYGMKLENVLFNMD 892
DB 946 KYEPTKSEKVEKLSSETNGSTNSTLEVPDVPQEKVAKFAESYGMKLENVLFNMD 1005
QY 893 GTIELYLPSPGEVTKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 952
DB 1006 GTIELYLPSPGEVTKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 1065
QY 953 PNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPADVPQOEKLEKFTASYGLGLDSV 1012
DB 1066 PNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPADVPQOEKLEKFTASYGLGLDSV 1125
QY 1013 IFNMDGTIELRPSGGEVIRKKNLSDEFTA 1039
DB 1126 IFNMDGTIELRPSGGEVIRKKNLSDEFTA 1152

RESULT 9

AAU84056
ID AAU84056 standard; Peptide; 1238 AA.
XX
AC AAU84056;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW29.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.

XX Streptococcus pneumoniae.
OS Synthetic.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
FT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia .
XX

XX Example 1; Page : 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 1238 AA;

Query Match 77.7%; Score 807; DB 23; Length 1238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NNQPSQSYSTASDNTQSVAKGSTSPANKSENLSLLKELYDSPSAQRYSESGLVLF 292
DB 432 NNQPSQSYSTASDNTQSVAKGSTSPANKSENLSLLKELYDSPSAQRYSESGLVLF 491
QY 293 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPIISGTGSTVSTNAKNEVVS 352

```
Db 492 DPAKIIISRTPNGVAIPHGDHYHFIPIYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 551
Qy 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 412
Db 552 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 611
Qy 413 TLPNNSLATPSPLPINPGTSHKHEEDGSGFDANRIIAEDESGFVMSHGDNHNYFFKDD 472
Db 612 TLPNNSLATPSPLPINPGTSHKHEEDGSGFDANRIIAEDESGFVMSHGDNHNYFFKDD 671
Qy 473 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 532
Db 672 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 731
Qy 533 KRESIVVKNKEKNAIYPHGDHHDADIDEHKPVGIGHSNHYELFPKEGVAKKEGNKYV 592
Db 732 KRESIVVKNKEKNAIYPHGDHHDADIDEHKPVGIGHSNHYELFPKEGVAKKEGNKYV 791
Qy 593 TGEELTNVNVLLKNSFTNNQFTLANGQKRVSPFPPELEKLGINMLVKLIITPDGKVLE 652
Db 792 TGEELTNVNVLLKNSFTNNQFTLANGQKRVSPFPPELEKLGINMLVKLIITPDGKVLE 851
Qy 653 KVSCKVFGEGVGNIANFELDPYLPQGTFKYTIASKDYPEVSDGTFVTPTSLAYKMASQ 712
Db 852 KVSCKVFGEGVGNIANFELDPYLPQGTFKYTIASKDYPEVSDGTFVTPTSLAYKMASQ 911
Qy 713 TIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGETIKLPIPKLNQ 772
Db 912 TIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGETIKLPIPKLNQ 971
Qy 773 TTRTAGNKIPVTMANAYLDNQSTYIVVEPILKEKNOTDKPSTLPQFKRKAQENSKLDE 832
Db 972 TTRTAGNKIPVTMANAYLDNQSTYIVVEPILKEKNOTDKPSTLPQFKRKAQENSKLDE 1031
Qy 833 KVEEPTSEKVEKEKLETSNGTSNSTLEEVPVTPVQVEKVAFAESYGMKLENVLNFKMD 892
Db 1032 KVEEPTSEKVEKEKLETSNGTSNSTLEEVPVTPVQVEKVAFAESYGMKLENVLNFKMD 1091
Qy 893 GTIELYLPSPGEVTKKNAADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPADSLPEA 952
Db 1092 GTIELYLPSPGEVTKKNAADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPADSLPEA 1151
Qy 953 PNEKPVKPNSTNGMLNPGNCGSDPMLDPALEEAAPVDPVQEKLEKFTASYGLGLDSV 1012
Db 1152 PNEKPVKPNSTNGMLNPGNCGSDPMLDPALEEAAPVDPVQEKLEKFTASYGLGLDSV 1211
Qy 1013 IFNMDGTIELRLPSGEVTKNLSDFIA 1039
Db 1212 IFNMDGTIELRLPSGEVTKNLSDFIA 1238

RESULT 10
AAU84057
ID AAU84057 standard; Peptide; 1365 AA.
XX
AC AAU84057;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW30.
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2..
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
```

```
XX
PR 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia.
XX
PS Example 1; Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g. caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocidia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1365 AA;
Query Match 77.7%; Score 807; DB 23; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 233 NMQPSQLSYSTASDNTQSVAKGSTSKPANKSENLOSLKELYDSPAQRYSDDGLVF 292
Db 559 NMQPSQLSYSTASDNTQSVAKGSTSKPANKSENLOSLKELYDSPAQRYSDDGLVF 618
Qy 293 DPAKIIISRTPNGVAIPHGDHYHFIPIYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 352
Db 619 DPAKIIISRTPNGVAIPHGDHYHFIPIYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 678
Qy 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 412
Db 679 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 738
Qy 413 TLPNNSLATPSPLPINPGTSHKHEEDGSGFDANRIIAEDESGFVMSHGDNHNYFFKDD 472
Db 739 TLPNNSLATPSPLPINPGTSHKHEEDGSGFDANRIIAEDESGFVMSHGDNHNYFFKDD 798
Qy 473 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 532
Db 799 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 858
Qy 533 KRESIVVKNKEKNAIYPHGDHHDADIDEHKPVGIGHSNHYELFPKEGVAKKEGNKYV 592
Db 859 KRESIVVKNKEKNAIYPHGDHHDADIDEHKPVGIGHSNHYELFPKEGVAKKEGNKYV 918
Qy 593 TGEELTNVNVLLKNSFTNNQFTLANGQKRVSPFPPELEKLGINMLVKLIITPDGKVLE 652
Db 919 TGEELTNVNVLLKNSFTNNQFTLANGQKRVSPFPPELEKLGINMLVKLIITPDGKVLE 978
```

QY 653 KVSQVGEVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKQASQ 712
 DB 979 KVSQVGEVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKQASQ 1038
 QY 713 TIFYPFHAGDYLVRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGEIKLPIPKLNQ 772
 DB 1039 TIFYPFHAGDYLVRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGEIKLPIPKLNQ 1098
 QY 773 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPQFKRKAQENSKLDE 832
 DB 1099 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPQFKRKAQENSKLDE 1158
 QY 833 KVEPKTSEKVEKEKLSGTNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 892
 DB 1159 KVEPKTSEKVEKEKLSGTNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 1218
 QY 893 GTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPADSLPEA 952
 DB 1219 GTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPADSLPEA 1278
 QY 953 PNEKPVKPSNTDNGMLNPEGVSGDPMLDPALEAFAPVDPQEKLEKFTASYGLGLDSV 1012
 DB 1279 PNEKPVKPSNTDNGMLNPEGVSGDPMLDPALEAFAPVDPQEKLEKFTASYGLGLDSV 1338
 QY 1013 IFNMDGTIELRLPDSGEVIRKKNLSDFIA 1039
 DB 1339 IFNMDGTIELRLPDSGEVIRKKNLSDFIA 1365

RESULT 11
 AAU84053
 ID AAU84053 standard; Peptide: 1378 AA.
 XX
 AC AAU84053;
 XX
 DT 08-MAY-2002 (first entry)
 DE S. pneumoniae derived chimeric peptide, NEW26.
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein;
 KW BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO2001198334-A2.
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX
 PS Example 1; Page -; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3. BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcadia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 1378 AA;

Query Match 77.7%; Score 807; DB 23; Length 1378;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENQSLKELYDSPAQRVSESGLVVF 292
 DB 572 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENQSLKELYDSPAQRVSESGLVVF 631
 QY 293 DPAKISRTNGVAIPHGDHYHFIPYKLSALEEKTARMVPISTGSTVSTNAKPNEVVS 352
 DB 632 DPAKISRTNGVAIPHGDHYHFIPYKLSALEEKTARMVPISTGSTVSTNAKPNEVVS 691
 QY 353 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIHPKSNIGQP 412
 DB 692 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIHPKSNIGQP 751
 QY 413 TLPNNSLATPSPSLPINPGTSHKHEEDGYGFANRTIAEDESSEGVMSHGDHNYHFEK 472
 DB 752 TLPNNSLATPSPSLPINPGTSHKHEEDGYGFANRTIAEDESSEGVMSHGDHNYHFEK 811
 QY 473 LTESQIKAAQKHLEEVKTSNGLDLSHSEQDYPGNAKEMKDLKDKTEEKIAGIMQYGV 532
 DB 812 LTESQIKAAQKHLEEVKTSNGLDLSHSEQDYPGNAKEMKDLKDKTEEKIAGIMQYGV 871
 QY 533 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKEEGVAKKEGKV 592
 DB 872 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKEEGVAKKEGKV 931
 QY 593 TGEELTNVWLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGINMLVKLITPDGKVL 652
 DB 932 TGEELTNVWLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGINMLVKLITPDGKVL 991
 QY 653 KVSQVGEVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKQASQ 712
 DB 992 KVSQVGEVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKQASQ 1051
 QY 713 TIFYPFHAGDYLVRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGEIKLPIPKLNQ 772
 DB 1052 TIFYPFHAGDYLVRVNPQFAVPKGTDALVRVDFEFGHNAVLENNYKVGEIKLPIPKLNQ 1111
 QY 773 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPQFKRKAQENSKLDE 832
 DB 1112 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPQFKRKAQENSKLDE 1171
 QY 833 KVEPKTSEKVEKEKLSGTNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 892
 DB 1172 KVEPKTSEKVEKEKLSGTNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 1231
 QY 893 GTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPADSLPEA 952
 DB 1232 GTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPADSLPEA 1291
 QY 953 PNEKPVKPSNTDNGMLNPEGVSGDPMLDPALEAFAPVDPQEKLEKFTASYGLGLDSV 1012
 DB 1291 PNEKPVKPSNTDNGMLNPEGVSGDPMLDPALEAFAPVDPQEKLEKFTASYGLGLDSV 1351

```
Db 1292 PNKPYKPNSTNDGMLNPEGNVSGMDPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1351
QY 1013 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
Db 1352 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1378

RESULT 12
AAB12744
ID AAB12744 standard; Protein; 780 AA.
AC AAB12744;
XX
DT 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae NEW15 protein antigen SEQ ID NO:78.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
PN WO200039299-A2.
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
PR 23-DEC-1998; 98US-0113800.
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Claim 18; Fig 43; 106pp; English.
XX
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW15 protein antigen.
XX
SQ Sequence 780 AA;

Query Match 75.1%; Score 780; DB 21; Length 780;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNHRSQENKNNRYSYVDGSSQSKSENLPDQVSQKEGTOAEQIVIKITDQGYV 80
Db 1 CAYALNHRSQENKNNRYSYVDGSSQSKSENLPDQVSQKEGTOAEQIVIKITDQGYV 60
QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGGYIITKVDGKYVYLKD 140
Db 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGGYIITKVDGKYVYLKD 120
QY 141 AAHADNVRTKDEINPKQEHVKDKNEKNSNVAVASQGRYTTNDGYVFNPAIDITGNA 200
Db 121 AAHADNVRTKDEINPKQEHVKDKNEKNSNVAVASQGRYTTNDGYVFNPAIDITGNA 180
QY 201 YTVPHGGHYHYTPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSTASDNNQTSVAKGSTSK 260
Db 181 YTVPHGGHYHYTPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSTASDNNQTSVAKGSTSK 240
```

```
QY 261 PANKSENQSLLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFIFYSK 320
Db 241 PANKSENQSLLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFIFYSK 300
QY 321 LSALEEKIARMVPISTGTSTVSTNAKPNNEVYSSLSGSSNPPSSLTTSKELSSASDGYIFN 380
Db 301 LSALEEKIARMVPISTGTSTVSTNAKPNNEVYSSLSGSSNPPSSLTTSKELSSASDGYIFN 360
QY 381 PKDIVEETATAYIVRHGDHPHYIPKSNQIQOPTLPNKSLATPSPSLPINPGTSHKHEED 440
Db 361 PKDIVEETATAYIVRHGDHPHYIPKSNQIQOPTLPNKSLATPSPSLPINPGTSHKHEED 420
QY 441 GYGF DANRIIAEDES GFVMSHGDNHYFFKKDLTEEQIKAAQKHLVEVKTSHRGLDSLSS 500
Db 421 GYGF DANRIIAEDES GFVMSHGDNHYFFKKDLTEEQIKAAQKHLVEVKTSHRGLDSLSS 480
QY 501 HEQDYPGNKEMKDLDDKIEKIIAGIMKQYGVKRESIVVANKENAIYIPHGDDHHADPID 560
Db 481 HEQDYPGNKEMKDLDDKIEKIIAGIMKQYGVKRESIVVANKENAIYIPHGDDHHADPID 540
QY 561 EHKPVGIGHSHSNYELFKPEGVAKEGKNVYTGEEELTNVNNLLKNSTFNQNFLLANGQ 620
Db 541 EHKPVGIGHSHSNYELFKPEGVAKEGKNVYTGEEELTNVNNLLKNSTFNQNFLLANGQ 600
QY 621 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVGKVGEGVGNIANFELDQPYLPQGT 680
Db 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVGKVGEGVGNIANFELDQPYLPQGT 660
QY 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQITIFYPFHAGDTYLRVNPQFAVFKGTDAL 740
Db 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQITIFYPFHAGDTYLRVNPQFAVFKGTDAL 720
QY 741 VRVDFEHGNAYLENNYKVGEEKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 800
Db 721 VRVDFEHGNAYLENNYKVGEEKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780

RESULT 13
AAB12748
ID AAB12748 standard; Protein; 1019 AA.
XX
AC AAB12748;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Disclosure; Fig 11; 106pp; English.
XX
```


CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 68.6%; Score 713; DB 21; Length 1019;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 21 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVDSQKEGIAQIVIKITDQGYV 80
DB 1 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVDSQKEGIAQIVIKITDQGYV 60

QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVKGYYIIKVDGKYVYVLD 140
DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVKGYYIIKVDGKYVYVLD 120

QY 141 AAHADNVRTKDEINRQKQEHVDKNEKVNNAVARSQGYTTNDGYVFNPAIIEDTGN 200
DB 121 AAHADNVRTKDEINRQKQEHVDKNEKVNNAVARSQGYTTNDGYVFNPAIIEDTGN 180

QY 201 YIVPHGGHYHYIPKSDLSASELAALAAHLAGKNMOPSOVSSTASDNNQTSVAKGTSK 260
DB 181 YIVPHGGHYHYIPKSDLSASELAALAAHLAGKNMOPSOVSSTASDNNQTSVAKGTSK 240

QY 261 PANKSENLOSLKELYDPSAORYSESGLVDPDAKIIISRTPNGVAIPHGDHYHFIYSK 320
DB 241 PANKSENLOSLKELYDPSAORYSESGLVDPDAKIIISRTPNGVAIPHGDHYHFIYSK 300

QY 321 LSALEKIAMVPIISGTSTVSTNAKPNEVSSLSLSPSSLTTSKELSSASDGYIFN 380
DB 301 LSALEKIAMVPIISGTSTVSTNAKPNEVSSLSLSPSSLTTSKELSSASDGYIFN 360

QY 381 PKDIVEETATAYIVRHGDHFHPIPSNQIQOPTLPNNSLATPSPLPIMGTSHEKHEED 440
DB 361 PKDIVEETATAYIVRHGDHFHPIPSNQIQOPTLPNNSLATPSPLPIMGTSHEKHEED 420

QY 441 GYGFDANRIIAEDSGFVMSHGDHNYFPFKDLTEEQKAAQKHLSEVKTSHGLDSSL 500
DB 421 GYGFDANRIIAEDSGFVMSHGDHNYFPFKDLTEEQKAAQKHLSEVKTSHGLDSSL 480

QY 501 HQDYPGNKAKMDLKKIEKIAIMKOYGVKRESIVVYNKEKNAILIYPHGDHHDADPID 560
DB 481 HQDYPGNKAKMDLKKIEKIAIMKOYGVKRESIVVYNKEKNAILIYPHGDHHDADPID 540

QY 561 EHKPVGIGHSHSNYELFKPEEGVAKGKGVYTGELTNVWLLKNSFNQNFILANGQ 620
DB 541 EHKPVGIGHSHSNYELFKPEEGVAKGKGVYTGELTNVWLLKNSFNQNFILANGQ 600

QY 621 KRVSFSPPELEKKGINLVLIITPDGKVLKGVKGVGEGVGNIANFELDQYLPQGT 680
DB 601 KRVSFSPPELEKKGINLVLIITPDGKVLKGVKGVGEGVGNIANFELDQYLPQGT 660

QY 681 FKYTIAKDYPEVSYDGTFTVPTSLAYKNASOTIFYPPHAGDTYLRVNPQFVAPKGTAL 740
DB 661 FKYTIAKDYPEVSYDGTFTVPTSLAYKNASOTIFYPPHAGDTYLRVNPQFVAPKGTAL 720

QY 741 VRVDFDHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPVTMANYLDNQSTIYIVE 800
DB 721 VRVDFDHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPVTMANYLDNQSTIYIVE 780

QY 801 VPILKENQTDPSILPQPKRNKAQNSKLDKVEBPKTSEKVEKEKLSSETNSNSTL 860
DB 781 VPILKENQTDPSILPQPKRNKAQNSKLDKVEBPKTSEKVEKEKLSSETNSNSTL 840

QY 861 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 920

DB 841 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 900
QY 921 GENKPSNGKYSTGTGVTENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPM 980
DB 901 GENKPSNGKYSTGTGVTENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPM 960
QY 981 LDPALERPAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKKNSLD 1036
DB 961 LDPALERPAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKKNSLD 1016

RESULT 14
AAB12752
ID AAB12752 standard; Protein; 1019 AA.
XX AAB12752;
XX AC
XX AC
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Disclosure; Fig 11: 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 68.6%; Score 713; DB 21; Length 1019;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 21 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVDSQKEGIAQIVIKITDQGYV 80
DB 1 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVDSQKEGIAQIVIKITDQGYV 60

QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVKGYYIIKVDGKYVYVLD 140
DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVKGYYIIKVDGKYVYVLD 120

QY 141 AAHADNVRTKDEINRQKQEHVDKNEKVNNAVARSQGYTTNDGYVFNPAIIEDTGN 200

Db	121	AAHADNVRTKDEINRQKQEHVKNEKVNVAVARSGRYTTNDGYVFNPAIIIDTGN	180
Qy	201	YIVPHGGHYHYIPKSDLSASELAHAKAGKNQPSQLSYSTASDNNTQSVAKGSTK	260
Db	181	YIVPHRGHYHYIPKSDLSASELAHAKAGKNQPSQLSYSTASDNNTQSVAKGSTK	240
Qy	261	PANKSENQSLLELYDSPAQRYSSESDGLVFPDAKIIISRTPNGVAIPHGDHYHFTPYSK	320
Db	241	PANKSENQSLLELYDSPAQRYSSESDGLVFPDAKIIISRTPNGVAIPHGDHYHFTPYSK	300
Qy	321	LSALEEKIARMVPISTGTVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	380
Db	301	LSALEEKIARMVPISTGTVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	360
Qy	381	PKDIVEETATAYIVRHGDHFFHYIPKSNQIGQPTLPNNSLATPSPSPINPCTSHEKHEED	440
Db	361	PKDIVEETATAYIVRHGDHFFHYIPKSNQIGQPTLPNNSLATPSPSPINPCTSHEKHEED	420
Qy	441	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSLS	500
Db	421	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSLS	480
Qy	501	HEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVKNKNAIYPHGDHHDADPID	560
Db	481	HEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVKNKNAIYPHGDHHDADPID	540
Qy	561	EHRPVGIGHSNVELFKPEGVAKGEGKNVYTGEEITNVNLLKSTNNQNTFLANGQ	620
Db	541	EHRPVGIGHSNVELFKPEGVAKGEGKNVYTGEEITNVNLLKSTNNQNTFLANGQ	600
Qy	621	KRVSFSPPELEKGLINMLVKLITPDGKVLKVSQKVFGEVGNIANFELDQYLPQGT	680
Db	601	KRVSFSPPELEKGLINMLVKLITPDGKVLKVSQKVFGEVGNIANFELDQYLPQGT	560
Qy	681	KYTIASKDYPEVSYDGTFTVPVSLAYKASQITFYFPHAGDTIYLRVNPQFAPVKGTDL	740
Db	661	KYTIASKDYPEVSYDGTFTVPVSLAYKASQITFYFPHAGDTIYLRVNPQFAPVKGTDL	720
Qy	741	VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLONQSYIYE	800
Db	721	VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLONQSYIYE	780
Qy	801	VPILEKENQTKPSIILPOFRKNAQNSKLDKVEBPCKSEKVEKEKLSGTNSNSTL	860
Db	781	VPILEKENQTKPSIILPOFRKNAQNSKLDKVEBPCKSEKVEKEKLSGTNSNSTL	840
Qy	861	EEVPTVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN	920
Db	841	EEVPTVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN	900
Qy	921	GENKPSNGKVTGTVENQNTENKPADSLPEAPNEKVPKPNSTDNQMLNPEGVGSQDM	980
Db	901	GENKPSNGKVTGTVENQNTENKPADSLPEAPNEKVPKPNSTDNQMLNPEGVGSQDM	960
Qy	981	LDPALEAPAVDPQVEKLEKFTASYGLGDSVLFNMDGTIELRPSGEVIKKNLSD	1036
Db	961	LDPALEAPAVDPQVEKLEKFTASYGLGDSVLFNMDGTIELRPSGEVIKKNLSD	1016
RESULT 15			
ID	AA12753	standard; Protein: 1019 AA.	
XX	AA12753;		
AC	AA12753;		
XX	21-NOV-2000	(first entry)	
DT	21-NOV-2000	(first entry)	
DE	Streptococcus pneumoniae strain A66 BVH-3	protein antigen.	
XX	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;		
KW	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;		
KW	otitis media; pneumonia; immunisation; bactericidal.		
XX			

OS Streptococcus pneumoniae.

XX WO200039299-A2.

XX 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 99US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N.

PI WPI: 2000-452397/39.

DR Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia

PS Disclosure; Fig 11; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the proteins

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents a

CC S. pneumoniae BVH-3 protein antigen, from the present invention.

XX SQ Sequence 1019 AA.

Query Match 68.6%; Score 713; DB 21; Length 1019;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps .0;

Qy 21 CAYALNQRHSEKDNRRVSYDGSQSKSENLTPOQVQKQEGIAQPIVIKITDQGV 80

Db 1 CAYALNQRHSEKDNRRVSYDGSQSKSENLTPOQVQKQEGIAQPIVIKITDQGV 60

Qy 81 TSHGDHYHYNCKVPYDALFSEELLMDPNYQLKADIVNEVKGYIHKVDGKYVYLKD 140

Db 61 TSHGDHYHYNCKVPYDALFSEELLMDPNYQLKADIVNEVKGYIHKVDGKYVYLKD 120

Qy 141 AAHADNVRTKDEINRQKQEHVKNEKVNVAVARSGRYTTNDGYVFNPAIIIDTGN 200

Db 121 AAHADNVRTKDEINRQKQEHVKNEKVNVAVARSGRYTTNDGYVFNPAIIIDTGN 180

Qy 201 YIVPHGGHYHYIPKSDLSASELAHAKAGKNQPSQLSYSTASDNNTQSVAKGSTK 260

Db 181 YIVPHRGHYHYIPKSDLSASELAHAKAGKNQPSQLSYSTASDNNTQSVAKGSTK 240

Qy 261 PANKSENQSLLELYDSPAQRYSSESDGLVFPDAKIIISRTPNGVAIPHGDHYHFTPYSK 320

Db 241 PANKSENQSLLELYDSPAQRYSSESDGLVFPDAKIIISRTPNGVAIPHGDHYHFTPYSK 300

Qy 321 LSALEEKIARMVPISTGTVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN 380

Db 301 LSALEEKIARMVPISTGTVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN 360

Qy 381 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGQPTLPNNSLATPSPSPINPCTSHEKHEED 440

Db 361 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGQPTLPNNSLATPSPSPINPCTSHEKHEED 420

Qy 441 GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSLS 500

Db 421 GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSLS 480

Qy 501 HEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVKNKNAIYPHGDHHDADPID 560

Db 481 HEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVKNKNAIYPHGDHHDADPID 540

561 EHKPVGIGHSHSNVELPKPEGVAKKGNKVTGEBELTNVLLKNSTFNQNFILANGQ 620
541 EHKPVGIGHSHSNVELPKPEGVAKKGNKVTGEBELTNVLLKNSTFNQNFILANGQ 600
621 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVSQVFGVGNIANFELDPYLPQGT 680
601 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVSQVFGVGNIANFELDPYLPQGT 660
681 EKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPFHAGDYLVRNPOFAPVPGTDL 740
661 EKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPFHAGDYLVRNPOFAPVPGTDL 720
741 VRVDFEHGNALENNYKVGIEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 800
721 VRVDFEHGNALENNYKVGIEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
801 VPILKENQTDKPSILPOFRNKAQENSKLDEKVEEPTSEKVEKEKLGSTGNSSTL 860
781 VPILKENQTDKPSILPOFRNKAQENSKLDEKVEEPTSEKVEKEKLGSTGNSSTL 840
861 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
841 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVKNSTNDGMLNPEGVNGSDPM 980
901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVKNSTNDGMLNPEGVNGSDPM 960
981 LDPALAEAPADVPQOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSD 1036
961 LDPALAEAPADVPQOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSD 1016

RESULT 16
AAU84092
ID AAU84092 standard; Peptide: 679 AA.
XX AC AAU84092;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW15.
XX KW BVH-3; BVH-11: vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; muten.
XX OS Streptococcus pneumoniae.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX PS WPI; 2002-122272/16.
XX DR New streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia -
XX PS Example 1; Page -; 113pp; English.
XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3. BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noardiae or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for S. pneumoniae infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological
sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating S. pneumoniae nucleic acid in a sample for
diagnosing streptococcal infections. This sequence represents a truncate
of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX
SQ Sequence 679 AA;
Query Match 65.4%; Score 679; DB 23; Length 679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 679; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CAYALNQHSQENKDNRYSDGSSQSSQSENLTDPQVSKQEGIAEQIVIKITDQGVY 80
DB 1 CAYALNQHSQENKDNRYSDGSSQSSQSENLTDPQVSKQEGIAEQIVIKITDQGVY 60
QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIIVNEVKGYYIIKVDGKYYVYLDK 140
DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIIVNEVKGYYIIKVDGKYYVYLDK 120
QY 141 AAHADNVRTKDEINRQKQEHVKNKVNNAVARSQGRYTTNDGYVFNPAIDTGN 200
DB 121 AAHADNVRTKDEINRQKQEHVKNKVNNAVARSQGRYTTNDGYVFNPAIDTGN 180
QY 201 YIVPHGCHYHYIPKSDLSASELAARAHLAGKNNQPSQLSYSTASDNNNTQSVAKGSTK 260
DB 181 YIVPHGCHYHYIPKSDLSASELAARAHLAGKNNQPSQLSYSTASDNNNTQSVAKGSTK 240
QY 261 PANKSENQLSKLKYDPSAQRYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 320
DB 241 PANKSENQLSKLKYDPSAQRYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
QY 321 LSALFEKIAEAPVPTSGTSTVSTNAKNEVYSSIGLSLSSNPSSLTTSKELSSASDGYIFN 380
DB 301 LSALFEKIAEAPVPTSGTSTVSTVSTNAKNEVYSSIGLSLSSNPSSLTTSKELSSASDGYIFN 360
QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 440
DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 441 GYGFDAIRITAEDESQVMSHGDHNYFFKKDLTEEQIKAAQKHLKLEVKTSNGLDLSLSS 500
DB 421 GYGFDAIRITAEDESQVMSHGDHNYFFKKDLTEEQIKAAQKHLKLEVKTSNGLDLSLSS 480
QY 501 HEQDYPGNAKEMKDLKKIEKIAKIMKOYGVKRESIVVANKERNAIYPHGDHHDADPID 560
DB 481 HEQDYPGNAKEMKDLKKIEKIAKIMKOYGVKRESIVVANKERNAIYPHGDHHDADPID 540
QY 561 EHKPVGIGHSHSNVELPKPEGVAKKGNKVTGEBELTNVLLKNSTFNQNFILANGQ 620
DB 541 EHKPVGIGHSHSNVELPKPEGVAKKGNKVTGEBELTNVLLKNSTFNQNFILANGQ 600
QY 621 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVSQVFGVGNIANFELDPYLPQGT 680
DB 601 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVSQVFGVGNIANFELDPYLPQGT 660
QY 681 EKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPFHAGDYLVRNPOFAPVPGTDL 740
DB 661 EKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPFHAGDYLVRNPOFAPVPGTDL 720

RESULT 17

AAU83996
ID AAU83996 standard; Peptide; 644 AA.

XX AC AAU83996;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-11, NEW21.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO2001198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX PT epitope-bearing polypeptides, useful as vaccine components for treating

XX PT or preventing streptococcal infections such as otitis media,

XX PT meningitis, and bacteraemia

XX PS Example 1; Page 7; 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an

XX CC individual susceptible to these disorders. (II) is also useful for

XX CC therapeutic or prophylactic treatment of any streptococcal bacterial

XX CC infection (e.g., caused by Streptococcus pneumoniae, group A

XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or

XX CC Staphylococcus aureus) in an individual susceptible to the infection.

XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic

XX CC test for S. pneumoniae infection. (III) is useful for designing DNA

XX CC probes for use in detecting the presence of Streptococcus in a biological

XX CC sample suspected of containing the bacteria. The DNA probes may also be

XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for

XX CC diagnosing streptococcal infections. This sequence represents a truncate

XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

XX CC described in the method of the invention.

XX CC Note: This sequence does not appear in the specification but has

XX CC been created according to information given in the invention.

XX XX Sequence 644 AA;

XX Query Match 62.0%; Score 644; DB 23; Length 644;

XX Best Local Similarity 100.0%; Pred. NO. 0;

XX Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 396 HCDHPHYTPKSNQIQPTLPNNLSATSPSPINPGTSHKHEEDGYGFDANRIIADES 455

XX

XX 1 HGDHPHYTPKSNQIQPTLPNNLSATSPSPINPGTSHKHEEDGYGFDANRIIADES 60

XX

XX 456 GFVMSHGDHNYFFKDLTEBQIKAAQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDL 515

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 61 GFVMSHGDHNYFFKDLTEBQIKAAQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDL 120
Qy 516 DKIEEKIAGIMKOYGVKRESIVVNVKKNKNAIIPYGHGHHADPIDHKHPVGIGHSNHYE 575
Db 121 DKIEEKIAGIMKOYGVKRESIVVNVKKNKNAIIPYGHGHHADPIDHKHPVGIGHSNHYE 180
Qy 576 LFKPEEGVAKKGNKYVTGEBELTNVNLNKNSTFNNOFTLANGOKRVKVSFSEPELEKLL 635
Db 181 LFKPEEGVAKKGNKYVTGEBELTNVNLNKNSTFNNOFTLANGOKRVKVSFSEPELEKLL 240
Qy 636 GINMLVLITPDGKVLKVSKEGEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSY 695
Db 241 GINMLVLITPDGKVLKVSKEGEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSY 300
Qy 696 DGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGNAYLEN 755
Db 301 DGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGNAYLEN 350
Qy 756 NYKGEIKLPIPKLNQGTTRTAGNKIPVTENANAYLDNQSTYIIVEVPILEKENQTKPSI 815
Db 361 NYKGEIKLPIPKLNQGTTRTAGNKIPVTENANAYLDNQSTYIIVEVPILEKENQTKPSI 420
Qy 815 LPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTNSSTLEEVPTVDVQEKVAK 875
Db 421 LPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTNSSTLEEVPTVDVQEKVAK 480
Qy 876 FAESYGMKLENVLFNMDGTIELPLPSGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGT 935
Db 481 FAESYGMKLENVLFNMDGTIELPLPSGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGT 540
Qy 936 VENOPTENKPADSLPEAPNEKPKPENSTNGMLNPEGNVGSDDPMLDPALEAPAVDPVQ 995
Db 541 VENOPTENKPADSLPEAPNEKPKPENSTNGMLNPEGNVGSDDPMLDPALEAPAVDPVQ 600
Qy 996 EKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 601 EKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 644

RESULT 18
AAU83996
ID AAU83996 standard; Peptide; 632 AA.
XX AC AAU83996;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-11, NEW40.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO2001198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX DR WPI; 2002-122272/16.
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

XX	Example 1: Page -: 113pp; English.	
PS	The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.	
CC	Note: This sequence does not appear in the specification but has been created according to information given in the invention.	
XX	Sequence 632 AA:	
SQ	Query Match 60.8%; Score 632; DB 23; Length 632; Best Local Similarity 100.0%; Pred. No. 0; Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	408 QIGQPTLPNNLSATPSLPINPGTSHKEHEEDGYCFDANRIIAEDSGFVMSHGDNHY 467	
DB	1 QIGQPTLPNNLSATPSLPINPGTSHKEHEEDGYCFDANRIIAEDSGFVMSHGDNHY 60	
QY	468 FFKKDLTEQIAAKQHLSEVTSINGLSDLSHSHQDTPGNAKEMKDLDKKIEEKIAGIM 527	
DB	61 FFKKDLTEQIAAKQHLSEVTSINGLSDLSHSHQDTPGNAKEMKDLDKKIEEKIAGIM 120	
QY	528 KQYGVKRESIVNKENAIYIPGHGHHADPTDEHKPVGIGHSHSNLYELFKPEEGVAKKE 587	
DB	121 KQYGVKRESIVNKENAIYIPGHGHHADPTDEHKPVGIGHSHSNLYELFKPEEGVAKKE 180	
QY	588 GKNVYTGEBLTNNVLLKNSTNNQNTLIANGOKRVSFSPPELEKKGINMLVLIITPD 647	
DB	181 GKNVYTGEBLTNNVLLKNSTNNQNTLIANGOKRVSFSPPELEKKGINMLVLIITPD 240	
QY	648 GKVLKGVSKGVGEGVNTANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLAY 707	
DB	241 GKVLKGVSKGVGEGVNTANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLAY 300	
QY	708 KMASQTFYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGHNAYLENNYKVGEEKLPI 767	
DB	301 KMASQTFYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGHNAYLENNYKVGEEKLPI 360	
QY	768 KLNQGTTRAGNKIPVTFMANAYLDNQSYIYEVPILEKENTDPSILPOPKRNKAQEN 827	
DB	361 KLNQGTTRAGNKIPVTFMANAYLDNQSYIYEVPILEKENTDPSILPOPKRNKAQEN 420	
QY	828 SKLDEKVEEPPKTSKYEKEKLSGTNSTNSLTLEEYPTVDPQEVAKFAESYGMKLENY 887	
DB	421 SKLDEKVEEPPKTSKYEKEKLSGTNSTNSLTLEEYPTVDPQEVAKFAESYGMKLENY 480	
QY	888 LFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSNGKSVTGTVENOPTENKPAD 947	
DB	481 LFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSNGKSVTGTVENOPTENKPAD 540	
QY	948 SLPEAPNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPAVDPQVEKLEKFTASYGL 1007	
DB	541 SLPEAPNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPAVDPQVEKLEKFTASYGL 600	
QY	1008 GLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039	

DB	601 GLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 632	
RESULT 19		
AAB12749		
ID	AAB12749 standard; Protein; 1019 AA.	
XX	AAB12749:	
DT	21-NOV-2000 (first entry)	
DE	Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.	
KW	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.	
XX	Streptococcus pneumoniae.	
PN	WO200039299-A2.	
PD	06-JUL-2000.	
PF	20-DEC-1999; 99WO-CA01218.	
PR	23-DEC-1998; 98US-0113800.	
XX	(BIOC-) BIOCHEM PHARMA INC.	
XX	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;	
DR	WPI: 2000-452397/39.	
PT	Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteraemia and/or pneumonia	
PS	Disclosure: Fig 11: 106pp; English.	
CC	The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents a S. pneumoniae BVH-3 protein antigen, from the present invention.	
SQ	Sequence 1019 AA:	
	Query Match 58.9%; Score 612; DB 21; Length 1019; Best Local Similarity 99.6%; Pred. No. 0; Matches 1012; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	21 CAYALNQHRSEKNNRVSYVDSQSSQKSENLTDPQVSKQEGIAQBIKIDQGYV 80	
DB	1 CAYALNQHRSEKNNRVSYVDSQSSQKSENLTDPQVSKQEGIAQBIKIDQGYV 60	
QY	81 TSHGDPHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGYIITKVDGKYYVILKD 140	
DB	61 TSHGDPHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGYIITKVDGKYYVILKD 120	
QY	141 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGN 200	
DB	121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGN 180	
QY	201 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAGKNMOPQSLSYSTASDNNTOSVAKGSTK 260	
DB	181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAGKNMOPQSLSYSTASDNNTOSVAKGSTK 240	
QY	261 PANKSENQSLKELYDPSAQRYSESDGLYDPAKIIISRTPNGVAIHGDHYHFIPIYSK 320	

Db 241 PANKSENUGSLKEIYDSPAQYSESGLVDFDPAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
Qy 321 LSALEKIAIRWPIISGTSTVSTNAKPNVSSLSGSLSNPSSLTTSKELSSASDGYIFN 380
Db 301 LSALEKIAIRWPIISGTSTVSTNAKPNVSSLSGSLSNPSSLTTSKELSSASDGYIFN 360
Qy 381 PKDIVEETATYIVRHGDHFIHPKSNQIGQPTLPNNSLATPSPLNPSTSHKHEED 440
Db 361 PKDIVEETATYIVRHGDHFIHPKSNQIGQPTLPNNSLATPSPLNPSTSHKHEED 420
Qy 441 GYGFDANRIIADESGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 500
Db 421 GYGFDANRIIADESGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 480
Qy 501 HEODYPGNAKEMKDLKTEEKIAGIMKOYGVKRESIVVYKKNALIIYPHGDHHDADPID 560
Db 481 HEODYPGNAKEMKDLKTEEKIAGIMKOYGVKRESIVVYKKNALIIYPHGDHHDADPID 540
Qy 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVYVLLKNSFNQNFPLANGQ 620
Db 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVYVLLKNSFNQNFPLANGQ 600
Qy 621 KRVSFPPPELEKKGILNMLVLIITPDGKLVKSGKVGEGVGNIANFELDPQYLPQOT 680
Db 601 KRVSFPPPELEKKGILNMLVLIITPDGKLVKSGKVGEGVGNIANFELDPQYLPQOT 660
Qy 681 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTAL 740
Db 661 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTAL 720
Qy 741 VRYDFEHGNAYLENNYKVGELKPLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYVE 800
Db 721 VRYDFEHGNAYLENNYKVGELKPLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYVE 780
Qy 801 VPILKENQDTPSILPQKRNKAQENSKLDEKVEEPTKSEKVEKEKISETGNSTNSL 860
Db 781 VPILKENQDTPSILPQKRNKAQENSKLDEKVEEPTKSEKVEKEKISETGNSTNSL 840
Qy 861 EEVPTVDPVQEKVAKPAESYGKMLNVLNMDGTIELYLPSEGEVKKKNMADFTGEAPQGN 920
Db 841 EEVPTVDPVQEKVAKPAESYGKMLNVLNMDGTIELYLPSEGEVKKKNMADFTGEAPQGN 900
Qy 921 GENKPSGENKVTGTVENOPTENKPADSLPEAPNEKPKPENSTNGMLNPGNVSDDPM 980
Db 901 GENKPSGENKVTGTVENOPTENKPADSLPEAPNEKPKPENSTNGMLNPGNVSDDPM 960
Qy 981 LDPALPEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSD 1036
Db 961 LDPALPEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSD 1016
RESULT 20
AAU84002
ID AAU84002 standard; Peptide; 632 AA.
AC AAU84002;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW49.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
XX Synthetic.
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX

PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WFI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Claim 1; Page : 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 632 AA;
Query Match 55.6%; Score 578; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 462 GDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEDQYPGNKEMKDLKKEE 521
Db 55 GDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEDQYPGNKEMKDLKKEE 114
Qy 522 KIAGIMKOYGVKRESIVVYKKNALIIYPHGDHHDADIDEKHPVIGHSHSNYELFKPEE 581
Db 115 KIAGIMKOYGVKRESIVVYKKNALIIYPHGDHHDADIDEKHPVIGHSHSNYELFKPEE 174
Qy 582 GVAKKEGKVVYTGELTNVYVLLKNSFNQNFPLANGQKRVSPSPPELEKILGINLV 641
Db 175 GVAKKEGKVVYTGELTNVYVLLKNSFNQNFPLANGQKRVSPSPPELEKILGINLV 234
Qy 642 KLITPDGKLVKSGKVGEGVGNIANFELDPQYLPQOTFKYTIASKDYPEVSDGTFTV 701
Db 235 KLITPDGKLVKSGKVGEGVGNIANFELDPQYLPQOTFKYTIASKDYPEVSDGTFTV 294
Qy 702 PTLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVG 761
Db 295 PTLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVG 354
Qy 762 IKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYVEPILEKENQTKPSILPQFKR 821
Db 355 IKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYVEPILEKENQTKPSILPQFKR 414
Qy 822 NKAQENSKLDEKVEEPTKSEKVEKEKISETGNSTNSLTVDPVQEKVAKFAESYG 881
Db 415 NKAQENSKLDEKVEEPTKSEKVEKEKISETGNSTNSLTVDPVQEKVAKFAESYG 474
Qy 882 MKLENVLFNMDGTIELYLPSEGEVKKKNMADFTGEAPQGNENKPSGVSTGTVENQPT 941

Db 475 MLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPT 534
Qy 942 ENKPADSLPEAPNEKPVKPNSTDMGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKF 1001
Db 535 ENKPADSLPEAPNEKPVKPNSTDMGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKF 594
Qy 1002 TASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 595 TASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 632

RESULT 21
AAU84003
ID AAU84003 standard; Peptide; 632 AA.
AC AAU84003;
XX
DT 08-MAY-2002 (first entry)
DE Truncated variant of S. pneumoniae BVH-3, NEW50.
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
PN 27-DEC-2001.
PD
PF 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
PR (SHIR-) SHIRE BIOCHEM INC.
PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI: 2002-122272/16.
DR
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX Claim 1; Page -: 113pp; English.
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX Sequence 632 AA;
SQ

Query Match 55.3%; Score 575; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 NHYFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIA 524
Db 58 NHYFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIA 117
Qy 525 GIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKEPVGIGHSHSNYELEKPEEGVA 584
Db 118 GIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKEPVGIGHSHSNYELEKPEEGVA 177
Qy 585 KKEGNKYVTGEELTNVYNLLKNSTFNQNTLANGQKRVSFSPPELEKLGINMLVKLI 644
Db 178 KKEGNKYVTGEELTNVYNLLKNSTFNQNTLANGQKRVSFSPPELEKLGINMLVKLI 237
Qy 645 TPDCKVLEKVGSKVFGGNGVNIANFELDQVLPQOTFKYTIASKDYPEVSYDGTFTVPTS 704
Db 238 TPDCKVLEKVGSKVFGGNGVNIANFELDQVLPQOTFKYTIASKDYPEVSYDGTFTVPTS 297
Qy 705 LAYKMASQTIIFYPEHAGDTYLVRNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGIEKL 764
Db 298 LAYKMASQTIIFYPEHAGDTYLVRNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGIEKL 357
Qy 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEYPILEKENQTDKPSILPOFKENKA 824
Db 358 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEYPILEKENQTDKPSILPOFKENKA 417
Qy 825 QENSKLDEKVEEPTSEKVEKELSETGNSTNSLTSEVPTVDPVQEKVAKFAESYGMKL 884
Db 418 QENSKLDEKVEEPTSEKVEKELSETGNSTNSLTSEVPTVDPVQEKVAKFAESYGMKL 477
Qy 885 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENK 944
Db 478 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENK 537
Qy 945 PADSLPEAPNEKPVKPNSTDMGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 538 PADSLPEAPNEKPVKPNSTDMGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKFTAS 597
Qy 1005 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 598 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 632

RESULT 22
AAU84013
ID AAU84013 standard; Peptide; 632 AA.
AC AAU84013;
XX
DT 08-MAY-2002 (first entry)
DE Truncated variant of S. pneumoniae BVH-3, NEW63.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
PN 27-DEC-2001.
PD
PF 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
PR (SHIR-) SHIRE BIOCHEM INC.
PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI

XX DB WPI; 2002-122272/16.

XX DB New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX DB epitope-bearing polypeptides, useful as vaccine components for treating

XX DB or preventing streptococcal infections such as otitis media,

XX DB meningitis, and bacteraemia

XX PS Claim 1; Page -: 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an

XX CC individual susceptible to these disorders. (II) is also useful for

XX CC therapeutic or prophylactic treatment of any streptococcal bacterial

XX CC infection (e.g., caused by Streptococcus pneumoniae, group A

XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novocardia or

XX CC Staphylococcus aureus) in an individual susceptible to the infection.

XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic

XX CC test for S. pneumoniae infection. (III) is useful for designing DNA

XX CC probes for use in detecting the presence of Streptococcus in a biological

XX CC sample suspected of containing the bacteria. The DNA probes may also be

XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for

XX CC diagnosing streptococcal infections. This sequence represents a truncate

XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

XX CC described in the method of the invention.

XX CC Note: This sequence does not appear in the specification but has

XX CC been created according to information given in the invention.

XX SX Sequence 632 AA;

Query Match 55.3%; Score 575; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEQIAAKAHLVEVKTSHNGDLSLSHEDYPCNAKEMKDLKTEEKIA 524

DB 58 NHYFFKDLTEQIAAKAHLVEVKTSHNGDLSLSHEDYPCNAKEMKDLKTEEKIA 117

QY 525 GIMKQYGVRESIVNKENALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEVA 584

DB 118 GIMKQYGVRESIVNKENALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEVA 177

QY 585 KKEGNKVTYGEELTNVNLKNTFNQNTLANGOKRVSFSPPELEKGLGNNLVKLI 644

DB 178 KKEGNKVTYGEELTNVNLKNTFNQNTLANGOKRVSFSPPELEKGLGNNLVKLI 237

QY 645 TPDGKVLKSVKVGEGVGNIANFELDQPLPGGTFKYTTASKDYPVSVXDGTFTVPTS 704

DB 238 TPDGKVLKSVKVGEGVGNIANFELDQPLPGGTFKYTTASKDYPVSVXDGTFTVPTS 297

QY 705 LAYKASQIIFYPFHAGDYLVRVQFAPVKCTDALVRVDFEFGHNAVLENNYKVGTEKL 764

DB 298 LAYKASQIIFYPFHAGDYLVRVQFAPVKCTDALVRVDFEFGHNAVLENNYKVGTEKL 357

QY 765 PIPKLNQSTRTAGNKIPVTFMANAYLDNQSTYIYVEVPLEKENQDTPKPSILPOFKRKA 824

DB 358 PIPKLNQSTRTAGNKIPVTFMANAYLDNQSTYIYVEVPLEKENQDTPKPSILPOFKRKA 417

QY 825 QNSKLDKEVEPKTSEKVEKLETSNSTLSNLTVEVPTVDPVQEKVAFASVGMKL 884

DB 418 QNSKLDKEVEPKTSEKVEKLETSNSTLSNLTVEVPTVDPVQEKVAFASVGMKL 477

QY 885 ENVLNMDGTIELYLPSEGVIKKKNMADTGEAPQNGENKPSSENGKVGSTGVENQPTENK 944

DB 478 ENVLNMDGTIELYLPSEGVIKKKNMADTGEAPQNGENKPSSENGKVGSTGVENQPTENK 537

QY 945 PADSLPEAPNEKPKPENSTNGMLNPGNVGSDPMLDPALEEAPVDPVQEKLEKFTAS 1004

DB 538 PADSLPEAPNEKPKPENSTNGMLNPGNVGSDPMLDPALEEAPVDPVQEKLEKFTAS 597

QY 1005 YGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039

DB 598 YGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 632

RESULT 23

AAU84018

ID AAU84018 standard; Peptide: 626 AA.

XX AC AAU84018;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, NEW105.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX PT epitope-bearing polypeptides, useful as vaccine components for treating

XX PT or preventing streptococcal infections such as otitis media,

XX PT meningitis, and bacteraemia

XX PS Claim 1; Page -: 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an

XX CC individual susceptible to these disorders. (II) is also useful for

XX CC therapeutic or prophylactic treatment of any streptococcal bacterial

XX CC infection (e.g., caused by Streptococcus pneumoniae, group A

XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novocardia or

XX CC Staphylococcus aureus) in an individual susceptible to the infection.

XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic

XX CC test for S. pneumoniae infection. (III) is useful for designing DNA

XX CC probes for use in detecting the presence of Streptococcus in a biological

XX CC sample suspected of containing the bacteria. The DNA probes may also be

XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for

XX CC diagnosing streptococcal infections. This sequence represents a truncate

XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

XX CC described in the method of the invention.

XX CC Note: This sequence does not appear in the specification but has

XX CC been created according to information given in the invention.

XX SX Sequence 626 AA;

Query Match 55.1%; Score 573; DB 23; Length 626;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
QY 467 YFFKDLTEEQIKAAQKHLVEVKTSHNGIDSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 526
Db 54 YFFKDLTEEQIKAAQKHLVEVKTSHNGIDSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 113
QY 527 MKOYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSHSNYELFKPEEGVAKK 586
Db 114 MKOYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSHSNYELFKPEEGVAKK 173
QY 587 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKKGILGIMLVKLITP 646
Db 174 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKKGILGIMLVKLITP 233
QY 647 DGKVLKESGKVGEGVGNIANFELDQYLPQGTQFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 234 DGKVLKESGKVGEGVGNIANFELDQYLPQGTQFKYTIASKDYPEVSYDGTFTVPTSLA 293
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFDEHGNAYLENNYKVGEEKLPI 766
Db 294 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFDEHGNAYLENNYKVGEEKLPI 353
QY 767 PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNQTDKPSILPQFRNKAAQE 826
Db 354 PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNQTDKPSILPQFRNKAAQE 413
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTVDPVQEKVAKFAESYGMKLEN 886
Db 414 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTVDPVQEKVAKFAESYGMKLEN 473
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPA 946
Db 474 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPA 533
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 534 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 593
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 594 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 626

RESULT 24
AAU84014
ID AAU84014 standard; Peptide; 632 AA.
XX AAU84014;
AC AAU84014;
XX
XX 08-MAY-2002 (first entry)
XX
XX Truncated variant of S. pneumoniae BVH-3, NEW64.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
```

```
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Claim 1: Page 7: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 632 AA;
```

Query Match 55.1%; Score 573; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 467 YFFKDLTEEQIKAAQKHLVEVKTSHNGIDSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 526
Db 60 YFFKDLTEEQIKAAQKHLVEVKTSHNGIDSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 119
QY 527 MKOYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSHSNYELFKPEEGVAKK 586
Db 120 MKOYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSHSNYELFKPEEGVAKK 179
QY 587 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKKGILGIMLVKLITP 646
Db 180 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKKGILGIMLVKLITP 239
QY 647 DGKVLKESGKVGEGVGNIANFELDQYLPQGTQFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 240 DGKVLKESGKVGEGVGNIANFELDQYLPQGTQFKYTIASKDYPEVSYDGTFTVPTSLA 299
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFDEHGNAYLENNYKVGEEKLPI 766
Db 300 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFDEHGNAYLENNYKVGEEKLPI 359
QY 767 PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNQTDKPSILPQFRNKAAQE 826
Db 360 PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNQTDKPSILPQFRNKAAQE 419
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTVDPVQEKVAKFAESYGMKLEN 886
Db 420 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTVDPVQEKVAKFAESYGMKLEN 479
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPA 946
Db 480 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPA 539
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 540 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 599
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 600 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 632
```

RESULT 25

AAU84070
ID AAU84070 standard; Peptide: 895 AA.

XX AC AAU84070;

XX DT 08-MAY-2002 (first entry)

XX DE S. pneumoniae derived chimeric peptide, VP112.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutain;
XX KW BVH-11-2.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

XX PS Claim 1; Page -; 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an
XX CC individual susceptible to these disorders. (II) is also useful for
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial

XX CC infection (e.g., caused by Streptococcus pneumoniae, group A
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
XX CC Staphylococcus aureus) in an individual susceptible to the infection.

XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic

XX CC test for S. pneumoniae infection. (III) is useful for designing DNA
XX CC probes for use in detecting the presence of Streptococcus in a biological

XX CC sample suspected of containing the bacteria. The DNA probes may also be
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for

XX CC diagnosing streptococcal infections. This sequence represents a chimeric
XX CC gene created from fragments and variant fragments of Streptococcus

XX CC pneumoniae genes, described in the method of the invention.

XX CC Note: This sequence does not appear in the specification but has

XX CC been created according to information given in the invention.

XX SQ Sequence 895 AA;

Query Match 55.1%; Score 573; DB 23; Length 895;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFPKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHDQDPYGNAKEMKDLDKKIEKIKGI 526
|||||

DB 323 YFFPKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHDQDPYGNAKEMKDLDKKIEKIKGI 382
|||||

QY 527 MKQYGVRESIVNKKENAIYPHGDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 566
|||||

DB 383 MKQYGVRESIVNKKENAIYPHGDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 442
|||||

QY 587 EGNKVTGEELTNVNNLLKNSTFNNQNFLLANGOKRVSFSPPELEKKGILNMLVLITP 646
|||||

DB 443 EGNKVTGEELTNVNNLLKNSTFNNQNFLLANGOKRVSFSPPELEKKGILNMLVLITP 502
|||||

QY 647 DGKYLEKSVKGVEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 706
|||||

DB 503 DGKYLEKSVKGVEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 562
|||||

QY 707 YKMASQIIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNKYGEIKLPI 766
|||||

DB 563 YKMASQIIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNKYGEIKLPI 622
|||||

QY 767 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 826
|||||

DB 523 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 682
|||||

QY 827 NSKLDERVEEPKTSKVEKEKLSGTGNTSNTSLSEVPTVDPVQEKVAKFAESYGHKLEN 886
|||||

DB 683 NSKLDERVEEPKTSKVEKEKLSGTGNTSNTSLSEVPTVDPVQEKVAKFAESYGHKLEN 742
|||||

QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPA 946
|||||

DB 743 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPA 802
|||||

QY 947 DSLPEAPNEKPKPENSTDNGLMNPENGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
|||||

DB 803 DSLPEAPNEKPKPENSTDNGLMNPENGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 862
|||||

QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
|||||

DB 863 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 895
|||||

RESULT 26
AAU84080
ID AAU84080 standard; Peptide: 896 AA.

XX AC AAU84080;

XX DT 08-MAY-2002 (first entry)

XX DE S. pneumoniae derived chimeric peptide, VP123.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutain;
XX KW BVH-11-2.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

```
PS Claim 1: Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 896 AA;

Query Match 55.1%; Score 573; DB 23; Length 896;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKENKDLKDKIEKIAGI 526
Db 324 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKENKDLKDKIEKIAGI 383

QY 527 MKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGICGSHSNYELFKPEGVAKK 586
Db 384 MKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGICGSHSNYELFKPEGVAKK 443

QY 587 EGNKYVTGELTNVNLKNSFNQNFANGQKRVSFSPPELEKLGINLWKLITP 646
Db 444 EGNKYVTGELTNVNLKNSFNQNFANGQKRVSFSPPELEKLGINLWKLITP 503

QY 647 DGKVLKSVKGVFEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGFTVPTSLA 706
Db 504 DGKVLKSVKGVFEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGFTVPTSLA 563

QY 707 YKMASOTIYFPFHAGDTYLRVNPQFAVPRGTDALVRVDFDHGNAYLENNYKVGELKPI 766
Db 564 YKMASOTIYFPFHAGDTYLRVNPQFAVPRGTDALVRVDFDHGNAYLENNYKVGELKPI 623

QY 767 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTDKPSILPQFKNKAGE 826
Db 624 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTDKPSILPQFKNKAGE 683

QY 827 NSKLDEKVEEPTSEKVEKEKLSNSETNSTLEEVTPDVPQEKVAFASVYGMKLEN 886
Db 684 NSKLDEKVEEPTSEKVEKEKLSNSETNSTLEEVTPDVPQEKVAFASVYGMKLEN 743

QY 887 VLFNMDGTIELYLPSEVTKKNNADFTGAPQCGNGENKPSNGKSVGTGVNQPTENKPA 946
Db 744 VLFNMDGTIELYLPSEVTKKNNADFTGAPQCGNGENKPSNGKSVGTGVNQPTENKPA 803

QY 947 DSLPEAPNEKPKVPENSTNGMLNPGNVGSDPMDLPALEAPAVDPVQEKLEKFTASVG 1006
Db 804 DSLPEAPNEKPKVPENSTNGMLNPGNVGSDPMDLPALEAPAVDPVQEKLEKFTASVG 863

QY 1007 LGLDSVIFNMDDGTIELRPSGVEIKKNSLDFIA 1039
Db 864 LGLDSVIFNMDDGTIELRPSGVEIKKNSLDFIA 896

RESULT 27
```

QY 587 EGNKYTGTELTNVNLLKNSFTNNQNTLANGOKRVSEFPPPELEKKGILGNMLVLTLP 646
 DB 449 EGNKYTGTELTNVNLLKNSFTNNQNTLANGOKRVSEFPPPELEKKGILGNMLVLTLP 508
 QY 647 DGKLVKSGKVFGEVGNIANFELDQYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLA 706
 DB 509 DGKLVKSGKVFGEVGNIANFELDQYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLA 568
 QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGELKLP 766
 DB 569 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGELKLP 628
 QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 826
 DB 629 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 688
 QY 827 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFSEYGMKLEN 886
 DB 689 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFSEYGMKLEN 748
 QY 887 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPA 946
 DB 749 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPA 808
 QY 947 DSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 1006
 DB 809 DSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 868
 QY 1007 LGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 1039
 DB 869 LGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 901

RESULT 28

AAU84072
 ID AAU84072 standard; Peptide; 901 AA.

XX AC AAU84072;
 XX DT 08-MAY-2002 (first entry)

XX DE S. pneumoniae derived chimeric peptide, vpl14.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 XX KW BVH-11-2.

XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX PN WO200198334-A2.
 XX PD 27-DEC-2001.
 XX PF 19-JUN-2001; 2001WO-CA00908.
 XX PR 20-JUN-2000; 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX DR WPI; 2002-122272/16.
 XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX PT epitope-bearing polypeptides, useful as vaccine components for treating
 XX PT or preventing streptococcal infections such as otitis media,
 XX PS meningitis, and bacteraemia
 XX PS Claim 1; Page -; 113pp; English.
 XX CC The invention describes an isolated polypeptide (1) with 70-90%

CC identity to streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 901 AA;

Query Match 55.1%; Score 573; DB 23; Length 901;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTETESQIAAKQHLLEEVKTSUNGSLSSSHEDQYPGNAKEMKDLCKIEEKIAGI 526
 DB 329 YFFKKDLTETESQIAAKQHLLEEVKTSUNGSLSSSHEDQYPGNAKEMKDLCKIEEKIAGI 388

QY 527 MKQYGVKRESIVVKNKKNALIIYPGHDDHDPIDEKPKVIGCHSHSNYELFPEEGVAKK 586
 DB 389 MKQYGVKRESIVVKNKKNALIIYPGHDDHDPIDEKPKVIGCHSHSNYELFPEEGVAKK 448

QY 587 EGNKYTGTELTNVNLLKNSFTNNQNTLANGOKRVSEFPPPELEKKGILGNMLVLTLP 646
 DB 449 EGNKYTGTELTNVNLLKNSFTNNQNTLANGOKRVSEFPPPELEKKGILGNMLVLTLP 508

QY 647 DGKLVKSGKVFGEVGNIANFELDQYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLA 706
 DB 509 DGKLVKSGKVFGEVGNIANFELDQYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLA 568

QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGELKLP 766
 DB 569 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGELKLP 628

QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 826
 DB 629 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 688

QY 827 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFSEYGMKLEN 886
 DB 689 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFSEYGMKLEN 748

QY 887 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPA 946
 DB 749 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPA 808

QY 947 DSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 1006
 DB 809 DSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 868

QY 1007 LGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 1039
 DB 869 LGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 901

RESULT 29

AAU84074

ID AAU84074 standard; Peptide; 901 AA.

XX XX

AC AAU84074;
XX
XX 08-MAY-2002 (first entry)
XX
XX S. pneumoniae derived chimeric peptide, VP116.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
XX Streptococcus pneumoniae.
OS Synthetic.
OS
XX WO200198334-A2.
XX
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
DR
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 901 AA;
SQ

Query Match
Best Local Similarity 55.1%; Score 573; DB 23; Length 901;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEEQIKAAQKHLVEVKTSHNGLDLSLSHEQYDPGNAKEMKDLDDKIEKIAGI 526
DB 329 YFFKKDLTEEQIKAAQKHLVEVKTSHNGLDLSLSHEQYDPGNAKEMKDLDDKIEKIAGI 388
QY 527 MKQYGVKRESIVVNEKNAIYPHGDHHPIDHKKVPGIGHSHSNVELEKPEEGVAKK 586
DB 389 MKQYGVKRESIVVNEKNAIYPHGDHHPIDHKKVPGIGHSHSNVELEKPEEGVAKK 448
QY 587 EGNKYVTGEELTNVYLLKNSTNNQNTLANGQKRVFSFPPPELEKKGINMLVKLITP 646

DB 449 EGNKYVTGEELTNVYLLKNSTNNQNTLANGQKRVFSFPPPELEKKGINMLVKLITP 508
QY 647 DGKLVKSVGKVFGEVGNIANFELDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLA 706
DB 509 DGKLVKSVGKVFGEVGNIANFELDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLA 568
QY 707 YKASQTIFFPFHAGDTYLRVNPQFAVPKGTDALRVVDFEHGHNAYLNNYKVEIKLPI 766
DB 569 YKASQTIFFPFHAGDTYLRVNPQFAVPKGTDALRVVDFEHGHNAYLNNYKVEIKLPI 628
QY 767 PKLNGQTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSILPQFKRNKAQE 826
DB 629 PKLNGQTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSILPQFKRNKAQE 688
QY 827 NSKLDERKEVPKTSKVEKEKLSGTNGSTNSLTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 689 NSKLDERKEVPKTSKVEKEKLSGTNGSTNSLTLEEVPTVDPVQEKVAKFAESYGMKLEN 748
QY 887 VLENMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPA 946
DB 749 VLENMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKSVSTGTVENOPTENKPA 808
QY 947 DSLPEAPNEKPVKPNSTDNMGMLNPEGNVSDPMDLPALEAPAVDPVQEKLEKFTASYG 1006
DB 809 DSLPEAPNEKPVKPNSTDNMGMLNPEGNVSDPMDLPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGLDSEVFNMDDGTIELRLPSEGEVIKKNLSDFIA 1039
DB 869 LGLDSEVFNMDDGTIELRLPSEGEVIKKNLSDFIA 901

RESULT 30
AAU84076
ID AAU84076 standard; Peptide; 902 AA.
XX
AC AAU84076;
XX
DT 08-MAY-2002 (first entry)
XX
XX S. pneumoniae derived chimeric peptide, VP119.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
DR
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia.
XX
XX Claim 1; Page : 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 901 AA;
SQ

CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (i) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX
 XX Sequence 902 AA;

Query Match 55.1%; Score 573; DB 23; Length 902;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIAKQKHLVEVKTSNGLDSSSHQDYPGNAKEMKDLKKIEKIAGI 526
 DB 330 YFFKDLTEQIAKQKHLVEVKTSNGLDSSSHQDYPGNAKEMKDLKKIEKIAGI 389

QY 527 MKQYGVKRESIVVKNKNAIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 586
 DB 390 MKQYGVKRESIVVKNKNAIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 449

QY 587 EGNKVTGTGELTNVNNLLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 646
 DB 450 EGNKVTGTGELTNVNNLLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 509

QY 647 DGKLVKSVGKVFGEVGNIAFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLA 706
 DB 510 DGKLVKSVGKVFGEVGNIAFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLA 569

QY 707 YKASOTIYPFHAGDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPI 766
 DB 570 YKASOTIYPFHAGDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPI 629

QY 767 PKLNOQTTRTAGNKIPVTMANAYLDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKQAE 826
 DB 630 PKLNOQTTRTAGNKIPVTMANAYLDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKQAE 689

QY 827 NSKLDKEVEPTKSEKVEKLEKSETGNSNSTLEVPDTPVQKVAKFAESYGMKLEN 886
 DB 690 NSKLDKEVEPTKSEKVEKLEKSETGNSNSTLEVPDTPVQKVAKFAESYGMKLEN 749

QY 887 VLFNMDGTTELPLPSGEVTKKNAADTGAPOGNGENKPSNGKYSTGTVENQPTENKPA 946
 DB 750 VLFNMDGTTELPLPSGEVTKKNAADTGAPOGNGENKPSNGKYSTGTVENQPTENKPA 809

QY 947 DSLPEAPNEKPKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPAVDPVOEKLKFTASYG 1006
 DB 810 DSLPEAPNEKPKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPAVDPVOEKLKFTASYG 869

QY 1007 LGLDSVIFNMDGTIELRLPSGEVTKKNSLDFIA 1039
 DB 870 LGLDSVIFNMDGTIELRLPSGEVTKKNSLDFIA 902

RESULT 31

AAU84078
 ID AAU84078 standard; Peptide; 902 AA.

XX
 AC AAU84078;

XX
 DT 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, VP121.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein;
 KW BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 OS WO200198334-A2.
 PN 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX Claim 1; Page : 113pp; English.
 PS The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

Sequence 902 AA;

Query Match 55.1%; Score 573; DB 23; Length 902;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIAKQKHLVEVKTSNGLDSSSHQDYPGNAKEMKDLKKIEKIAGI 526
 DB 330 YFFKDLTEQIAKQKHLVEVKTSNGLDSSSHQDYPGNAKEMKDLKKIEKIAGI 389

QY 527 MKQYGVKRESIVVKNKNAIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 586
 DB 390 MKQYGVKRESIVVKNKNAIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 449

QY 587 EGNKVTGTGELTNVNNLLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 646
 DB 450 EGNKVTGTGELTNVNNLLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 509

QY 647 DGKLVKSVGKVFGEVGNIAFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLA 706

Db 510 DGVLEKVSQVFGVGVGNIAFELDQPYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLA 569
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
Db 570 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 629
QY 767 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILILEKENOTDKPSILPOFKRNKAQE 826
Db 630 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILILEKENOTDKPSILPOFKRNKAQE 689
QY 827 NSKLDKVEEPTSEKVEKEKLETSNSTLSNSTLEEVPTVDPVQEKVAFASYNKLEN 886
Db 690 NKLDEKVEEPTSEKVEKEKLETSNSTLSNSTLEEVPTVDPVQEKVAFASYNKLEN 749
QY 887 VLFNMDGTIELPSGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGVNQPTENKPA 946
Db 750 VLFNMDGTIELPSGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGVNQPTENKPA 809
QY 947 DSLPEAPNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPAYDPVQEKLEKFTASYG 1006
Db 810 DSLPEAPNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPAYDPVQEKLEKFTASYG 869
QY 1007 LGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 870 LGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 902
RESULT 32
AAB12731
ID AAB12731 standard; Protein; 568 AA.
XX AAB12731;
XX AAB12731;
DT 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
DE Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW Proprietary; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX WO200039299-A2.
XX 06-JUL-2000.
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX Claim 18; Fig 29; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW1 protein antigen.

SQ Sequence 568 AA;
Query Match 54.7%; Score 568; DB 21; Length 568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEEQKAAQKHLEEVKTSNGLSLSSHEQDYPGNAKEMKDLCKIEEKIAGIMQYQ 531
Db 1 DLTEEQKAAQKHLEEVKTSNGLSLSSHEQDYPGNAKEMKDLCKIEEKIAGIMQYQ 60
QY 532 VKRESIVVYKNEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFPPEGVAKKEGKV 591
Db 61 VKRESIVVYKNEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFPPEGVAKKEGKV 120
QY 592 YTGELTNVWLLNSTFNQNFNLQNGKRVSEFPELEKLGINMLVKLIIPDGKVL 651
Db 121 YTGELTNVWLLNSTFNQNFNLQNGKRVSEFPELEKLGINMLVKLIIPDGKVL 180
QY 652 EKVSQVFGVGVGNIAFELDQPYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
Db 181 EKVSQVFGVGVGNIAFELDQPYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 240
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIKLNQ 771
Db 241 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIKLNQ 300
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILILEKENOTDKPSILPOFKRNKAENSKLD 831
Db 301 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILILEKENOTDKPSILPOFKRNKAENSKLD 360
QY 832 EKVEEPTSEKVEKEKLETSNSTLSNSTLEEVPTVDPVQEKVAFASYNKLENVLFNM 891
Db 361 EKVEEPTSEKVEKEKLETSNSTLSNSTLEEVPTVDPVQEKVAFASYNKLENVLFNM 420
QY 892 DGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGVNQPTENKPADSLPE 951
Db 421 DGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGVNQPTENKPADSLPE 480
QY 952 APNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPAYDPVQEKLEKFTASYGLGLDS 1011
Db 481 APNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPAYDPVQEKLEKFTASYGLGLDS 540
QY 1012 VTFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 541 VTFNMDGTIELRLPSGEVIKKNLSDFIA 568
RESULT 33
AAU83999
ID AAU83999 standard; Peptide; 568 AA.
XX AAU83999;
XX AAU83999;
DT 08-MAY-2002 (first entry)
DE Truncated variant of S. pneumoniae BVH-3, NEW1.
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX

```
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Claim 1: Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
XX Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 568 AA:
XX
XX Query Match 54.7%; Score 568; DB 23; Length 568;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 472 DLPEEQTKAAQKHLKLEVKTSNGLSLSSHEQDYPGNKEMKDLKKIEKIAGIMKQYV 531
XX 1 DLPEEQTKAAQKHLKLEVKTSNGLSLSSHEQDYPGNKEMKDLKKIEKIAGIMKQYV 60
XX
XX QY 532 VKRESIVNKEKNAIIPGHGHHADPIDHKPVGIGHSNHYELFKPEGVAKKGNKV 591
XX 61 VKRESIVNKEKNAIIPGHGHHADPIDHKPVGIGHSNHYELFKPEGVAKKGNKV 120
XX
XX QY 592 YTGELTNVNLKSTFNQNTLANGKRVSFSPFPPPELEKLGINMLVKLITPDGKYL 651
XX 121 YTGELTNVNLKSTFNQNTLANGKRVSFSPFPPPELEKLGINMLVKLITPDGKYL 180
XX
XX QY 652 EKVSQVFGGVGNINAFELDQPLPGQTKFTIASKDYPVSVYDGTFTVPSLAYKMAS 711
XX 181 EKVSQVFGGVGNINAFELDQPLPGQTKFTIASKDYPVSVYDGTFTVPSLAYKMAS 240
XX
XX QY 712 QTIYFPFHAGTTLRVNPOFVAPKGTDALRVFDFEFGHAYLNNYKVGKELPKLPKLNQ 771
XX 241 QTIYFPFHAGTTLRVNPOFVAPKGTDALRVFDFEFGHAYLNNYKVGKELPKLPKLNQ 300
XX
XX QY 772 GTTTRAGNKIPVTFMANAYLDNSTIVVEVPILEKENQTDKPSILPQFRNKAQENSKLD 831
XX 301 GTTTRAGNKIPVTFMANAYLDNSTIVVEVPILEKENQTDKPSILPQFRNKAQENSKLD 360
XX
XX QY 832 EKVEEPTSKVEKELSETGNSTNSTLEEVPTDVPQVKFAESYGMKLENVLFNM 891
XX 361 EKVEEPTSKVEKELSETGNSTNSTLEEVPTDVPQVKFAESYGMKLENVLFNM 420
XX
XX QY 892 DGTIELYLPSPSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 951
XX 421 DGTIELYLPSPSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 480
XX
XX QY 952 APNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
XX
XX Db 481 APNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 540
XX
XX QY 1012 VIFNMDGTIELRLPSPGEVKKNLSDRIA 1039
XX 541 VIFNMDGTIELRLPSPGEVKKNLSDRIA 568
XX
XX RESULT 34
XX AAU84059
XX ID AAU84059 standard; Peptide: 913 AA.
XX
XX AC AAU84059;
XX
XX DT 08-MAY-2002 (first entry)
XX
XX DE S. pneumoniae derived chimeric peptide, NEW32.
XX
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutain;
XX BVH-11-2.
XX
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX
XX PN W0200198334-A2.
XX
XX PD 27-DEC-2001.
XX
XX PF 19-JUN-2001; 2001WO-CA00908.
XX
XX PR 20-JUN-2000; 2000US-212683P.
XX
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Example 1: Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
XX Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 913 AA:
XX
XX Query Match 54.7%; Score 568; DB 23; Length 913;
XX Best Local Similarity 100.0%; Pred. No. 0;
```


QY 1012 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
 |||||
 Db 972 VIFNMDGTIELRLPSGEVIKKNLSDFIA 999

RESULT 36

AAU84052
ID AAU84052 standard. Peptide: 999 AA.

AC AAU84052:

08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, NEW20.

AA
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; murein;
KW BVH-11-2.

OS Streptococcus pneumoniae.

OS Synthetic.

PN WO200198334-A2.

PD 27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

New *Streptococcus pneumoniae* BVH-3 and BVH-11 variant and peptide-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia -

XX Example 1; Page -: 113pp; English. PS

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novarum or S. saprophyticus), or in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Query Match 54.7%: Score 568: DB 23: Length 999:

Query Match
Best Local Similarity
100.0%; Pred. No. 0;
Score 300;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 472·DLTEEQIKAAQKHLEEVKTSHNGLDLSLSHEQDYPGNAKEMKDLDKKIEEKIAGIMKQYG 531

Db	2	DUTEQIKAAQKHLEEVKTSNGLDLSLSSHQDYPGNAKENMKDLDKIEEKIAGIMKQYG	61
QY	532	VKRESIVNKEKNAIYIPHGDSHHADP IDEHKPVGIGHSHSNYELFPBEGVAKGKNKV	591
Db	62	VKRESIVNKEKNAIYIPHGDSHHADP IDEHKPVGIGHSHSNYELFPBEGVAKGKNKV	121
QY	592	YTGEEELTNVYNLLKNSTNNQNF LANGOKRVSVSFPPELEKKLGINMLVKLIIPDGKVL	651
Db	122	YTGEEELTNVYNLLKNSTNNQNF LANGOKRVSVSFPPELEKKLGINMLVKLIIPDGKVL	181
QY	652	EKVSGRVFGEGVGNIANFELDOPYLPGOTFKYTTASKDYPEVSYDGTFTVPTSLAYKMAS	711
Db	182	EKVSGRVFGEGVGNIANFELDOPYLPGOTFKYTTASKDYPEVSYDGTFTVPTSLAYKMAS	241
QY	712	QITIFYPFHAGDYTLRVNPOFAVPKGTDLRVFDFEFGHNAYLENNYKVGEEKLPIPKLNQ	771
Db	242	QITIFYPFHAGDYTLRVNPOFAVPKGTDLRVFDFEFGHNAYLENNYKVGEEKLPIPKLNQ	301
QY	772	GTTTRTAGNKIPVTENANAYLDNQSYIYVEVPTLEKENOTDKPSTILPOFKRNKAQENSKLD	831
Db	302	GTTTRTAGNKIPVTENANAYLDNQSYIYVEVPTLEKENOTDKPSTILPOFKRNKAQENSKLD	361
QY	832	EKVEEPEKTSKYEKEKLSGTNSTLEEYPTVDPQEKVAFASGYMKLENVLFNM	891
Db	362	EKVEEPEKTSKYEKEKLSGTNSTLEEYPTVDPQEKVAFASGYMKLENVLFNM	421
QY	892	DGTIELYLPBSGEVIKKNMADFTGEAPOGNGENKPSKNGKSTGTGVENQPTENKFADSLPE	951
Db	422	DGTIELYLPBSGEVIKKNMADFTGEAPOGNGENKPSKNGKSTGTGVENQPTENKFADSLPE	481
QY	952	APNEKPVKPNSTNDGMLNPEGVNGSDPMLDPALPEAPVDPVOEKLKFTFASYGGLDS	1011
Db	482	APNEKPVKPNSTNDGMLNPEGVNGSDPMLDPALPEAPVDPVOEKLKFTFASYGGLDS	541
QY	1012	VIFNMDGTTIELRLPSGEVIKKNLSDFIA	1039
Db	542	VIFNMDGTTIELRLPSGEVIKKNLSDFIA	569

37 RESULT.

AAB12725

AA12725
ID AA12725 standard; Protein; 1057 AA.

AC AAB12725:

XX	DT	21-NOV-2000	(first entry)

streptococcus pneumoniae NEW12 protein antigen SEQ ID NO:58.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

XX WO200039299-A2

XX
PD 06-JUL-2000.XX 20-DEC-1999: 99WO-CA01218.
PF

XX 23-DEC-1998: 98US-0113800.

XX PA (BTCC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N; PT

XX
DR WPI: 2000-452397/39.

AA Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteremia and/or pneumonia -
PT

Qy

PS

Claim 18; Fig 23; 106pp; English.

XX

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae NEW12 protein antigen.

XX

SQ

Sequence 1057 AA;

Query Match 54.7%; Score 568; DB 21; Length 1057;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

472 DITEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEEKIAGIMKQYG 531

DB

1 DITEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEEKIAGIMKQYG 60

QY

532 VKRESIVVNKEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591

DB

61 VKRESIVVNKEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 120

QY

592 YNGEELTNVNLKKNSTFNNQFTLANGOKRVSPFPPELEKKGINMLVLIPTDQKVL 651

DB

121 YNGEELTNVNLKKNSTFNNQFTLANGOKRVSPFPPELEKKGINMLVLIPTDQKVL 180

QY

652 ERVSGKVFEGVGNIANFELDPQYLPFGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711

DB

181 ERVSGKVFEGVGNIANFELDPQYLPFGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 240

QY

712 QTIFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 771

DB

241 QTIFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 300

QY

772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 831

DB

301 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 360

QY

832 EKVEPKTSEKVEKELSTGTSTNSLSEEVPTVDPVOEKVAKFAESYGMKLENVLNFM 891

DB

361 EKVEPKTSEKVEKELSTGTSTNSLSEEVPTVDPVOEKVAKFAESYGMKLENVLNFM 420

QY

892 DGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSNGKVGSTGTVENOPTENKPADSLPE 951

DB

421 DGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSNGKVGSTGTVENOPTENKPADSLPE 480

QY

952 APNEKVPKPNSTDNMLNPEGNVGSDPMLDPALEAPAVDPVQKLEKFTASYGLGLDS 1011

DB

481 APNEKVPKPNSTDNMLNPEGNVGSDPMLDPALEAPAVDPVQKLEKFTASYGLGLDS 540

QY

1012 VIFNMDGTIELPLSPGEVIVKKNLSDFIA 1039

DB

541 VIFNMDGTIELPLSPGEVIVKKNLSDFIA 568

RESULT 38

ID

AAU84097

XX

AAU84097 standard; Peptide; 1058 AA.

AC

AAU84097;

DT

08-MAY-2002 (first entry)

XX

S. pneumoniae derived chimeric peptide, NEW 17.

DE

XX

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW

pneumonia; streptococcal bacterial infection; mutant; mutein;

KW

BVH-11-2.

XX

OS

Streptococcus pneumoniae.

XX

PN

WO200198334-A2.

XX

PD

27-DEC-2001.

XX

PF

19-JUN-2001; 2001WO-CA00908.

XX

PR

20-JUN-2000; 2000US-212683P.

XX

PA

(SHIR-) SHIRE BIOCHEM INC.

XX

PI

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX

DR

WPI; 2002-122272/16.

XX

PT

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT

epitope-bearing polypeptides, useful as vaccine components for treating

PT

or preventing streptococcal infections such as otitis media,

PT

meningitis, and bacteraemia

XX

PS

Example 1; Page -: 113pp; English.

XX

CC

The invention describes an isolated polypeptide (I) with 70-90%

CC

identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC

BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC

comprising (I) is useful for therapeutic or prophylactic treatment of

CC

meningitis, otitis media, bacteraemia or pneumonia infection in an

CC

individual susceptible to these disorders. (II) is also useful for

CC

therapeutic or prophylactic treatment of any streptococcal bacterial

CC

infection (e.g., caused by Streptococcus pneumoniae, group A

CC

Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC

as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or

CC

Staphylococcus aureus) in an individual susceptible to the infection.

CC

A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC

techniques. The Streptococcus polypeptides are useful in a diagnostic

CC

test for S. pneumoniae infection. (III) is useful for designing DNA

CC

probes for use in detecting the presence of Streptococcus in a biological

CC

sample suspected of containing the bacteria. The DNA probes may also be

CC

used for detecting circulating S. pneumonia nucleic acid in a sample for

CC

diagnosing streptococcal infections. This sequence represents a chimeric

CC

gene created from fragments and variant fragments of Streptococcus

CC

pneumoniae genes, described in the method of the invention.

CC

Note: This sequence does not appear in the specification but has

CC

been created according to information given in the invention.

XX

SQ

Sequence 1058 AA;

Query Match 54.7%; Score 568; DB 23; Length 1058;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

472 DITEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEEKIAGIMKQYG 531

DB

2 DITEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEEKIAGIMKQYG 61

QY

532 VKRESIVVNKEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591

DB

62 VKRESIVVNKEKNAIIPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 121

QY

592 YNGEELTNVNLKKNSTFNNQFTLANGOKRVSPFPPELEKKGINMLVLIPTDQKVL 651

DB

122 YNGEELTNVNLKKNSTFNNQFTLANGOKRVSPFPPELEKKGINMLVLIPTDQKVL 181

QY

652 EKVSKEGVEGEGVGNIANFELDPQYLPFGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711

DB

182 EKVSKEGVEGEGVGNIANFELDPQYLPFGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 241

QY

712 QTIFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 771

DB

242 QTIFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 301

QY

772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 831

Db	302	GTTRTAGNKIPVTFMANAYLDNQSTIVVEPILEKENQTDKPSILPQKRKAQENSKLD	361
Qy	832	EKVESEPKTSEKVEKELSETGNSNSTLEEVPVDVQEKVAKFAESYGMKLENVLFNM	891
Db	362	EKVESEPKTSEKVEKELSETGNSNSTLEEVPVDVQEKVAKFAESYGMKLENVLFNM	421
Qy	892	DGTIELYLPSCGEVIKKKNMADFTGAPOGNGENKPSNGSKYSTGTVENQPTENKPADSLPE	951
Db	422	DGTIELYLPSCGEVIKKKNMADFTGAPOGNGENKPSNGSKYSTGTVENQPTENKPADSLPE	481
Qy	952	ANEXPKVPENSTNGMLNPGNVGSDPMLDDPALEAPAVDPVQEKLEKFTASYGLGLDS	1011
Db	482	ANEXPKVPENSTNGMLNPGNVGSDPMLDDPALEAPAVDPVQEKLEKFTASYGLGLDS	541
Qy	1012	VIFNMDGTIELRLPSGEVIKKKNLSDFIA	1039
Db	542	VIFNMDGTIELRLPSGEVIKKKNLSDFIA	569
RESULT 39			
AAU84058			
ID	AAU84058	standard; Peptide; 1126 AA.	
XX	AAU84058;		
XX	AC		
XX	AC		
DT	08-MAY-2002	(first entry)	
XX		S. pneumoniae derived chimeric peptide, NEW31.	
DE		BVH-3; BVH-11; vaccine; meningitis; Otitis media; bacteraemia;	
XX		pneumonia; streptococcal bacterial infection; mutant; mutein;	
KW		BVH-11-2.	
KW			
XX		Streptococcus pneumoniae.	
OS		Synthetic.	
XX		WO200198334-A2.	
PN		27-DEC-2001.	
XX		19-JUN-2001; 2001WO-CA00908.	
PF		20-JUN-2000; 2000US-212683P.	
XX		(SHIR-) SHIRE BIOCHEM INC.	
PA		Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
PI		WPI; 2002-122272/16.	
XX		New Streptococcus pneumoniae BVH-3 and BVH-11 variant and	
PT		epitope-bearing polypeptides, useful as vaccine components for treating	
PT		or preventing streptococcal infections such as otitis media,	
PT		meningitis, and bacteraemia	
XX		Example 1; Page -: 113pp; English.	
XX		The invention describes an isolated polypeptide (I) with 70-90%	
CC		identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of	
CC		BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)	
CC		comprising (I) is useful for therapeutic or prophylactic treatment of	
CC		meningitis, otitis media, bacteraemia or pneumonia infection in an	
CC		individual susceptible to these disorders. (II) is also useful for	
CC		therapeutic or prophylactic treatment of any streptococcal bacterial	
CC		infection (e.g., caused by Streptococcus pneumoniae, group A	
CC		Streptococcus such as Streptococcus pyogenes, group B Streptococcus such	
CC		as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or	
CC		Staphylococcus aureus) in an individual susceptible to the infection.	
CC		A polynucleotide (III) encoding (I) is useful in DNA immunisation	
CC		techniques. The Streptococcus polypeptides are useful in a diagnostic	
CC		test for S. pneumoniae infection. (III) is useful for designing DNA	
CC		probes for use in detecting the presence of Streptococcus in a biological	

CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of *Streptococcus*
CC *pneumoniae* genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX
SQ Sequence 1126 AA;

Query Match 54.7%; Score 568; DB 23; Length 1126;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQJKAQKHLEEVKTSHINGLDSLSSSHEDQYPGNAKEMKDKDKKIEEIGIMQYQ 531
DB 559 DLTEQJKAQKHLEEVKTSHINGLDSLSSSHEDQYPGNAKEMKDKDKKIEEIGIMQYQ 618
QY 532 VKRESIVVNKEKNAIYPGHGDHHDADIDEHKPYGIGHSHSNYELFPEEGVAKKEGNKV 591
DB 619 VKRESIVVNKEKNAIYPGHGDHHDADIDEHKPYGIGHSHSNYELFPEEGVAKKEGNKV 678
QY 592 YTGELTNVNVLLKNSFNNOFTLANGQKRVSFSPPELEKKLGINMLVLIITPDGKVL 651
DB 679 YTGELTNVNVLLKNSFNNOFTLANGQKRVSFSPPELEKKLGINMLVLIITPDGKVL 738
QY 652 EKVGSKVFGCGVGNIANFELDQYLPQGTFRKYTIASKDYPVSYDGTFTVPTSLAYKMAS 711
DB 739 EKVGSKVFGCGVGNIANFELDQYLPQGTFRKYTIASKDYPVSYDGTFTVPTSLAYKMAS 798
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALRVDFEFGHGNAYLENNYKVGEEKLPIPKLNQ 771
DB 799 QTIFYPFHAGDTYLRVNPQFAVPKGTDALRVDFEFGHGNAYLENNYKVGEEKLPIPKLNQ 858
QY 772 GTTRTAGNKIPVTMANAYLONQSTYIIVEVPILKEKENTDKPSILPQFKRKAQENSKLD 831
DB 859 GTTRTAGNKIPVTMANAYLONQSTYIIVEVPILKEKENTDKPSILPQFKRKAQENSKLD 918
QY 832 EKVEEPTKSEKVEKEKISGTGNSNSTNLTLEEVPTVDPQVKVAKFAESYGMKLENVLNFM 891
DB 919 EKVEEPTKSEKVEKEKISGTGNSNSTNLTLEEVPTVDPQVKVAKFAESYGMKLENVLNFM 978
QY 892 DGTIELYLPSEVGIKKNAADFTGEAPQNGENKPKSENGKVGSTGVENOPTENKPADSLPE 951
DB 979 DGTIELYLPSEVGIKKNAADFTGEAPQNGENKPKSENGKVGSTGVENOPTENKPADSLPE 1038
QY 952 APNEKPKVPKPNSTDNGLNPNCGNSVDGPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 1039 APNEKPKVPKPNSTDNGLNPNCGNSVDGPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1098
QY 1012 VIFNMDGTIELPLSGEVIKKNLSDFIA 1039
DB 1099 VIFNMDGTIELPLSGEVIKKNLSDFIA 1126

RESULT 40
RAU84055
ID AAU84055 standard; Peptide: 1139 AA.
XX AAU84055;
XX AAU84055;
XX AAU84055;
XX AAU84055;
DT 08-MAY-2002 (first entry)
XX
DE *S. pneumoniae* derived chimeric peptide, NEW28.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; muten;
KW BVH-11-2.
XX
OS *Streptococcus pneumoniae*.
OS Synthetic.
OS
PN WO200198334-A2.

```

XX 27-DEC-2001.
PD
PF
XX 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1139 AA;
Query Match 54.7%; Score 568; DB 23; Length 1139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKRIEIKIAGIMKQYG 531
DB 572 DLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKRIEIKIAGIMKQYG 631
QY 532 VKRESIVVNEKNALIIYPGHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
DB 632 VKRESIVVNEKNALIIYPGHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 691
QY 592 YTGELTNVLLKNSTNNONFTLANGQKRVSFSPPELEKLGINMLVKLIIPDGKVL 651
DB 692 YTGELTNVLLKNSTNNONFTLANGQKRVSFSPPELEKLGINMLVKLIIPDGKVL 751
QY 652 EKVSQKVEGNGVNIANFELQDPLPGQTFYITASKDYPVSDGTFVTPTSILAYKMAS 711
DB 752 EKVSQKVEGNGVNIANFELQDPLPGQTFYITASKDYPVSDGTFVTPTSILAYKMAS 811
QY 712 QTIFYPFAGTYLRVNPQFAPVPGTDLVRVDFEFGHNAVLENKYKGEIKLPIKLQ 771
DB 812 QTIFYPFAGTYLRVNPQFAPVPGTDLVRVDFEFGHNAVLENKYKGEIKLPIKLQ 871
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPFILEKENQTDKPSILPOFKRNKAQENSKLD 831
DB 872 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPFILEKENQTDKPSILPOFKRNKAQENSKLD 931

```

RESULT 41

```

AAU84006
ID AAU84006 standard; Peptide; 632 AA.
XX
AC AAU84006;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW53.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
XX
OS Synthetic.
XX
XX WO200198334-A2.
PN
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

```

CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX
SQ Sequence 632 AA;

Query Match 51.1%; Score 531; DB 23; Length 632;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 631; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 408 QIQGPTLPNNSLATPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSGHDHNY 467
DB 1 QIQGPTLPNNSLATPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSGHDHNY 60
QY 468 FFKKOLTEQIKAAQHLEEVKTSINGLSDSSSHEDQYPGNAKEMKLDKKIEKIKAGIM 527
DB 61 FFKKOLTEQIKAAQHLEEVKTSINGLSDSSSHEDQYPGNAKEMKLDKKIEKIKAGIM 120
QY 528 KOYGKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSHSNYELFKPEEGVAKKE 587
DB 121 KOYGKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSHSNYELFKPEEGVAKKE 180
QY 588 GNKVTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPELEKLGIMLVKLTIPD 647
DB 181 GNKVTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPELEKLGIMLVKLTIPD 240
QY 648 GKVLKESGKVGEGVGNIANFELDQPLPGQTKYTIASKDYPEVSDGTFTVPTSLAY 707
DB 241 GKVLKESGKVGEGVGNIANFELDQPLPGQTKYTIASKDYPEVSDGTFTVPTSLAY 300
QY 708 KMASOTIYFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHGNAYLENNYKVGELKIP 767
DB 301 KMASOTIYFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHGNAYLENNYKVGELKIP 360
QY 768 KLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTKPSILPOFKRNKAQEN 827
DB 361 KLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTKPSILPOFKRNKAQEN 420
QY 828 SKLDEKVEEPTSEKVEKEKLETSNSTLEEVPTVDPVQEKVAKFAESYGMKLENV 887
DB 421 SKLDEKVEEPTSEKVEKEKLETSNSTLEEVPTVDPVQEKVAKFAESYGMKLENV 480
QY 888 LFNMDGTIELYLPSEGVTKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPAD 947
DB 481 LFNMDGTIELYLPSEGVTKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPAD 540
QY 948 SLPEAPNEKPVKPNSTNGMLNPGNVSDDPMLDPALEAPAVDPVQEKLEKFTASYGL 1007
DB 541 SLPEAPNEKPVKPNSTNGMLNPGNVSDDPMLDPALEAPAVDPVQEKLEKFTASYGL 600
QY 1008 GLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 1039
DB 601 GLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 632

RESULT 42

AAB12719

ID AAB12719 standard; Protein: 528 AA.

XX

AC AAB12719;

XX

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

XX

OS Streptococcus pneumoniae.

XX

PN W0200039299-A2.

XX

PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
(BIOC-) BIOCHEM PHARMA INC.
XX
PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
PI WPI; 2000-452397/39.
XX
DR N-PSDB; AAA65734.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Clalm 18; Fig 10; 106pp; English.
XX
CC The present invention describes nucleic acids (i) encoding protein
CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3B protein antigen.
XX
SQ Sequence 528 AA;

Query Match 50.8%; Score 528; DB 21; Length 528;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 MKDLDDKIEEKIAGIMKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 571
DB 1 MKDLDDKIEEKIAGIMKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 60
QY 572 SNLEFKPEEGVAKKGNKYVTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPEL 631
DB 61 SNLEFKPEEGVAKKGNKYVTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPEL 120
QY 632 EKKLGIMLVKLTIPDGKYLEKVGKVGEGVGNIANFELDQPLPGQTKYTIASKDYP 691
DB 121 EKKLGIMLVKLTIPDGKYLEKVGKVGEGVGNIANFELDQPLPGQTKYTIASKDYP 180
QY 692 EVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGH 751
DB 181 EVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGH 240
QY 752 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTD 811
DB 241 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTD 300
QY 812 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLETSNSTLEEVPTVDPVQEK 871
DB 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLETSNSTLEEVPTVDPVQEK 360
QY 872 KVAFAESYGMKLENVLFNMDGTIELYLPSEGVTKKNMADFTGEAPQNGENKPSNGKV 931
DB 361 KVAFAESYGMKLENVLFNMDGTIELYLPSEGVTKKNMADFTGEAPQNGENKPSNGKV 420
QY 932 STGTGVNQPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNVSDDPMLDPALEAPAV 991
DB 421 STGTGVNQPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNVSDDPMLDPALEAPAV 480
QY 992 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVKKNLSDFTA 1039
DB 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVKKNLSDFTA 528

RESULT 43

AAB64024

ID AU84024 standard; Peptide; 528 AA.
 XX AC AU84024;
 XX DT 08-MAY-2002 (first entry)
 XX DE Truncated variant of S. pneumoniae BVH-3, BVH-3B.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutin.
 XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX PN WO2001198334-A2.
 XX PD 27-DEC-2001.
 XX PF 19-JUN-2001; 2001WO-CA00908.
 XX PR 20-JUN-2000; 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.
 XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX DR WPI; 2002-122272/16.
 XX PS Example 1; Page -; 113pp; English.
 XX CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX SQ Sequence 528 AA;
 XX Query Match 50.8%; Score 528; DB 23; Length 528;
 XX Best Local Similarity 100.0%; Pred. No. 0;
 XX Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 512 MKDLKKIEETAGIMKQYGVKRESIVVYKNEKNAIIPHGHHADPDIEHKPVGIGHSH 571
 DB 1 MKDLKKIEETAGIMKQYGVKRESIVVYKNEKNAIIPHGHHADPDIEHKPVGIGHSH 60
 QY 572 SNYELFKPEEGVAKREGNKVITGEELTNVLLKNSTNNONFTLANGKQKVSFPPPEL 631
 DB 61 SNYELFKPEEGVAKREGNKVITGEELTNVLLKNSTNNONFTLANGKQKVSFPPPEL 120
 QY 632 EKKLGINMLVKLITPDGKVLKSVKGVGEGVGNIANFELDQPYLPQCTFKYTIASKDYP 691

Db 121 EKKLGINMLVKLITPDGKVLKSVKGVGEGVGNIANFELDQPYLPQCTFKYTIASKDYP 180
 QY 692 EVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 751
 Db 181 EVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
 QY 752 YLENNYKVGELKPIPKLNOGTTTACNKPVTTFMANAYLDNQSTIYVEVPILEKENQTD 811
 Db 241 YLENNYKVGELKPIPKLNOGTTTACNKPVTTFMANAYLDNQSTIYVEVPILEKENQTD 300
 QY 812 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKEKLESETGNSNSTLEEVPTVDPVQE 871
 Db 301 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKEKLESETGNSNSTLEEVPTVDPVQE 360
 QY 872 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKMADFTGEAPQNGENKPSENGKV 931
 Db 361 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKMADFTGEAPQNGENKPSENGKV 420
 QY 932 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGDPMLEDPALAEAPAV 991
 Db 421 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGDPMLEDPALAEAPAV 480
 QY 992 DPVOEKLEKFTASYGLGDSVIFNMDGTIELRLPSPGSEVIKKNLSDFIA 1039
 Db 481 DPVOEKLEKFTASYGLGDSVIFNMDGTIELRLPSPGSEVIKKNLSDFIA 528
 RESULT 44
 AAB12724
 ID AAB12724 standard; Protein; 509 AA.
 XX AC AAB12724;
 XX DT 21-NOV-2000 (first entry)
 XX DE Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
 XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX KW otitis media; pneumonia; immunisation; bactericidal.
 XX OS Streptococcus pneumoniae.
 XX PN WO200039299-A2.
 XX PD 06-JUL-2000.
 XX PF 20-DEC-1999; 99WO-CA01218.
 XX PR 23-DEC-1998; 98US-0113800.
 XX PA (BIOC-) BIOCHEM PHARMA INC.
 XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX DR WPI; 2000-452397/39.
 XX ST Streptococcal antigens useful for vaccinating against e.g. meningitis,
 XX PT otitis media, bacteraemia and/or pneumonia -
 XX PS Claim 18; Fig 22; 106pp; English.
 XX CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae L-BVH-3-AD protein antigen.

```

SQ      Sequence      509 AA;
Query Match      49.0%; Score 509; DB 21; Length 509;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MKFSKKYIAAGSAVIVSLCAVALNQHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60
Db      1 MKFSKKYIAAGSAVIVSLCAVALNQHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60
Qy      61 QKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALFSELLMKDPNQLKDADIVN 120
Db      61 QKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALFSELLMKDPNQLKDADIVN 120
Qy      121 EVKGGYIIKVDGKYVYVLKDAHAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGRY 180
Db      121 EVKGGYIIKVDGKYVYVLKDAHAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGRY 180
Qy      181 TTNDGYVFNPAADIETDGNAYIVPHGGHYHYIPKSDLSASELAHAHLAGKNMOPSQLS 240
Db      181 TTNDGYVFNPAADIETDGNAYIVPHGGHYHYIPKSDLSASELAHAHLAGKNMOPSQLS 240
Qy      241 YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSESGLVFDPAKIISR 300
Db      241 YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSESGLVFDPAKIISR 300
Qy      301 TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVYSSLSLSSN 360
Db      301 TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVYSSLSLSSN 360
Qy      361 PSSLTTSKELSSASDGYIENPKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLA 420
Db      361 PSSLTTSKELSSASDGYIENPKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLA 420
Qy      421 TPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEEQIKA 480
Db      421 TPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEEQIKA 480
Qy      481 AQKHLEEVKTSHNGLDLSLSSHEQDYPGNA 509
Db      481 AQKHLEEVKTSHNGLDLSLSSHEQDYPGNA 509

RESULT 45
AAU84023
ID      AAU84023 standard; Peptide; 509 AA.
XX      AC      AAU84023;
XX      DT      08-MAY-2002 (first entry)
XX      DE      Truncated variant of S. pneumoniae BVH-3, L-BVH-3AD.
KW      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW      pneumonia; streptococcal bacterial infection; mutant; mutenb.
XX      OS      Streptococcus pneumoniae.
XX      PN      WO200198334-A2.
XX      PD      27-DEC-2001.
XX      PR      19-JUN-2001; 2001WO-CA00908.
XX      PR      20-JUN-2000; 2000US-212683P.
XX      PA      (SHIR-) SHIRE BIOCHEM INC.
XX      PI      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      DR      WPI; 2002-122272/16.
XX

```

```

PT      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
PT      meningitis, and bacteraemia
XX      Example 1; Page -: 113pp; English.
XX      The invention describes an isolated polypeptide (I) with 70-90%
CC      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
CC      individual susceptible to these disorders. (II) is also useful for
CC      therapeutic or prophylactic treatment of any streptococcal bacterial
CC      infection (e.g., caused by Streptococcus pneumoniae, group A
CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC      Streptococcus aureus) in an individual susceptible to the infection.
CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC      techniques. The Streptococcus polypeptides are useful in a diagnostic
CC      test for S. pneumoniae infection. (III) is useful for designing DNA
CC      probes for use in detecting the presence of Streptococcus in a biological
CC      sample suspected of containing the bacteria. The DNA probes may also be
CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
CC      diagnosing streptococcal infections. This sequence represents a truncate
CC      of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC      described in the method of the invention.
CC      Note: This sequence does not appear in the specification but has
CC      been created according to information given in the invention.
XX      SQ      Sequence      509 AA;
Query Match      49.0%; Score 509; DB 23; Length 509;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MKFSKKYIAAGSAVIVSLCAVALNQHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60
Db      1 MKFSKKYIAAGSAVIVSLCAVALNQHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60
Qy      61 QKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALFSELLMKDPNQLKDADIVN 120
Db      61 QKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALFSELLMKDPNQLKDADIVN 120
Qy      121 EVKGGYIIKVDGKYVYVLKDAHAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGRY 180
Db      121 EVKGGYIIKVDGKYVYVLKDAHAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGRY 180
Qy      181 TTNDGYVFNPAADIETDGNAYIVPHGGHYHYIPKSDLSASELAHAHLAGKNMOPSQLS 240
Db      181 TTNDGYVFNPAADIETDGNAYIVPHGGHYHYIPKSDLSASELAHAHLAGKNMOPSQLS 240
Qy      241 YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSESGLVFDPAKIISR 300
Db      241 YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSESGLVFDPAKIISR 300
Qy      301 TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVYSSLSLSSN 360
Db      301 TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVYSSLSLSSN 360
Qy      361 PSSLTTSKELSSASDGYIENPKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLA 420
Db      361 PSSLTTSKELSSASDGYIENPKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLA 420
Qy      421 TPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEEQIKA 480
Db      421 TPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEEQIKA 480
Qy      481 AQKHLEEVKTSHNGLDLSLSSHEQDYPGNA 509
Db      481 AQKHLEEVKTSHNGLDLSLSSHEQDYPGNA 509

```


Search completed: May 13, 2003, 14:00:04
Job time : 66 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:58:38 ; Search time 26 Seconds
(without alignments)
3841.679 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKKYIAAGSAVIVSLSL.....TELRLPGEVTKNLSDFIA 1039

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 45

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : PIR-73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	834	80.3	1039	2 H95115	conserved hypothe
2	632	60.8	1039	2 D97985	hypotheical prote
3	32	3.1	802	2 G95136	conserved domain p
4	32	3.1	819	2 B95136	conserved domain p
5	32	3.1	828	2 E98004	hypotheical prote
6	32	3.1	839	2 G95115	conserved hypothe
7	32	3.1	853	2 C97985	hypotheical prote
8	32	3.1	855	2 D98004	histidine Motif-O
9	16	1.5	822	2 T46758	hypotheical 92.4K
10	9	0.9	45	2 E97985	hypotheical prote
11	9	0.9	182	2 F97985	hypotheical prote
12	8	0.8	92	2 C69541	conserved hypothe
13	8	0.8	96	2 A61291	ferredoxin [2Fe-2S
14	8	0.8	97	1 FEED	ferredoxin [2Fe-2S
15	8	0.8	97	1 FEYB6	ferredoxin [2Fe-2S
16	8	0.8	127	2 S39689	ywd protein - Bac
17	8	0.8	139	1 PER2	ferredoxin [2Fe-2S
18	8	0.8	146	2 A96919	PTS system, mannit
19	8	0.8	154	2 AD3475	protein-Npi-phosph
20	8	0.8	207	2 A69349	conserved hypothe
21	8	0.8	233	1 R5HG1T	ribosomal protein
22	8	0.8	273	1 R5R22	ribosomal protein
23	8	0.8	273	1 R5M2	ribosomal protein
24	8	0.8	275	2 T11810	ribosomal protein
25	8	0.8	275	2 T07361	ribosomal protein
26	8	0.8	276	2 T07361	ribosomal protein
27	8	0.8	277	2 AF2987	ABC transporter, m
28	8	0.8	278	2 B98296	sulfate ABC transp
29	8	0.8	278	2 A13633	hypotheical prote

30 8 0.8 327 2 H59094
31 8 0.8 345.2 2 G83148
32 8 0.8 410 2 AD2087
33 8 0.8 472 2 F81329
34 8 0.8 522 2 B83735
35 8 0.8 573 2 B97331
36 8 0.8 672 2 S61157
37 8 0.8 753 2 S48261
38 8 0.8 763 2 T18280
39 8 0.8 765 2 T40674
40 8 0.8 813 2 G96494
41 8 0.8 840 2 C75313
42 8 0.8 984 2 G86393
43 8 0.8 1148 2 S72635
44 8 0.8 1218 2 T14265
45 8 0.8 1237 2 T45070

ALIGNMENTS

RESULT 1

H95115
Conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <NR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1004

Query Match 80.3%; Score 834; DB 2; Length 1039;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1034; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MKFSKKYIAAGSAVIVSLSLCAVALNQHRQENKNNRVSYVDGSSQSKSENLTDPQVS 60
Db 1 MKFSKKYIAAGSAVIVSLSLCAVALNQHRQENKNNRVSYVDGSSQSKSENLTDPQVS 60
Qy 61 QKGIQAEQIVIKITDQGYVTSHGDIHYNGKVPYDALFSELLMKDPNYQLKDADIVN 120
Db 61 QKGIQAEQIVIKITDQGYVTSHGDIHYNGKVPYDALFSELLMKDPNYQLKDADIVN 120
Qy 121 EVKGGYIIVKVDGKYVYLKDAHADNVRTKDEINRQKQEHVKNKVNNAVARSQRY 180
Db 121 EVKGGYIIVKVDGKYVYLKDAHADNVRTKDEINRQKQEHVKNKVNNAVARSQRY 180
Qy 181 TTNDGVYFNPAIDETGNAYIVPHGGHYHYIPKSDLSASELAAXAHLAGKNQFSQLS 240
Db 181 TTNDGVYFNPAIDETGNAYIVPHGGHYHYIPKSDLSASELAAXAHLAGKNQFSQLS 240
Qy 241 YSTASDNNTQSAVGSTSKPANKSENQLQLKELXDSQAQRYSDGLVFPDAKIIISR 300
Db 241 YSTASDNNTQSAVGSTSKPANKSENQLQLKELXDSQAQRYSDGLVFPDAKIIISR 300
Qy 301 TPNGVAIPHGDHYHFFPYSKLSALEEKIARMVPISGTGTSTVSTNAKPNVWSGLSLSSN 360
Db 301 TPNGVAIPHGDHYHFFPYSKLSALEEKIARMVPISGTGTSTVSTNAKPNVWSGLSLSSN 360
Qy 361 PSSLTTSKELSSASDGIYFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQPTLPNNSLA 420

Db 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRGDHFHIFPKSNQIGQPTLPNLSA 420
QY 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNRHYFFKDLTEEQIKA 480
Db 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNRHYFFKDLTEEQIKA 480
QY 481 AOKHLEEVKTSNGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
Db 481 AOKHLEEVKTSNGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
QY 541 KEKNAIIPHGDRHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
Db 541 KEKNAIIPHGDRHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
QY 601 VNLKSTFNQNTLANGOKRVSPFPPELEKLGINMLVLIITPDGKVLKESGVKVF 660
Db 601 VNLKSTFNQNTLANGOKRVSPFPPELEKLGINMLVLIITPDGKVLKESGVKVF 660
QY 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
Db 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
QY 721 GDTYLRVNPQFVAPKCTDALRVDFDEHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
Db 721 GDTYLRVNPQFVAPKCTDALRVDFDEHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
QY 781 IPVTFMANAYLDNQSTYIVVEPILKENTDKPSILPQKRNKAQENKLDKEVEEPKTS 840
Db 781 IPVTFMANAYLDNQSTYIVVEPILKENTDKPSILPQKRNKAQENKLDKEVEEPKTS 840
QY 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLENLNFNMDGTIELYLP 900
Db 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLENLNFNMDGTIELYLP 900
QY 901 SGEVYKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKVPKP 960
Db 901 SGEVYKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKVPKP 960
QY 961 ENSTDNGMLNPEGVNSDPMLEAPFAVDPVQEKLEKFTASYGLGLDSVIFNMDGTI 1020
Db 961 ENSTDNGMLNPEGVNSDPMLEAPFAVDPVQEKLEKFTASYGLGLDSVIFNMDGTI 1020
QY 1021 ELRLPSGEVIKKNLSD 1036
Db 1021 ELRLPSGEVIKKNLSD 1036

RESULT 2
D97985
hypoetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUP>
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:gl5458515; GSPDB:GN00174
C:Gene: phtE

Query Match 60.8%; Score 632; DB 2; Length 1039;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1032; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKFSKKYIAAGSAVIVSLCAAYALNQHRSQENKNNRVSYVDGSSQSKSENITPDQVS 60

Db 1 MKFSKKYIAAGSAVIVSLCAAYALNQHRSQENKNNRVSYVDGSSQSKSENITPDQVS 60
QY 61 QKEGIAEQIVIKITDQGYVTSGHDHYHYNGKPYDADFSEELMKDPNYPQLKADIVN 120
Db 61 QKEGIAEQIVIKITDQGYVTSGHDHYHYNGKPYDADFSEELMKDPNYPQLKADIVN 120
QY 121 EVKGGYIIVKVDGKYVYLKDAAHADNRVTKEINRQKQEHYKDNKVNNAVARSQGR 180
Db 121 EVKGGYIIVKVDGKYVYLKDAAHADNRVTKEINRQKQEHYKDNKVNNAVARSQGR 180
QY 181 TTNDGVFNPAADIIETGNAYIVPHGGHYHYIPKSDLSASELAALAAKAHLAGKNQPSOLS 240
Db 181 TTNDGVFNPAADIIETGNAYIVPHGGHYHYIPKSDLSASELAALAAKAHLAGKNQPSOLS 240
QY 241 YSSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPSAQRYSESGLVDFDAKIISR 300
Db 241 YSSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPSAQRYSESGLVDFDAKIISR 300
QY 301 TPNGVAIPHGDRHVFIPYKSLSALEEKIARMVPISGTGSTVSTNAKPNVYSSLSGSSN 360
Db 301 TPNGVAIPHGDRHVFIPYKSLSALEEKIARMVPISGTGSTVSTNAKPNVYSSLSGSSN 360
QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRGDHFHIFPKSNQIGQPTLPNLSA 420
Db 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRGDHFHIFPKSNQIGQPTLPNLSA 420
QY 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNRHYFFKDLTEEQIKA 480
Db 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNRHYFFKDLTEEQIKA 480
QY 481 AOKHLEEVKTSNGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
Db 481 AOKHLEEVKTSNGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
QY 541 KEKNAIIPHGDRHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
Db 541 KEKNAIIPHGDRHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
QY 601 VNLKSTFNQNTLANGOKRVSPFPPELEKLGINMLVLIITPDGKVLKESGVKVF 660
Db 601 VNLKSTFNQNTLANGOKRVSPFPPELEKLGINMLVLIITPDGKVLKESGVKVF 660
QY 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
Db 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
QY 721 GDTYLRVNPQFVAPKCTDALRVDFDEHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
Db 721 GDTYLRVNPQFVAPKCTDALRVDFDEHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
QY 781 IPVTFMANAYLDNQSTYIVVEPILKENTDKPSILPQKRNKAQENKLDKEVEEPKTS 840
Db 781 IPVTFMANAYLDNQSTYIVVEPILKENTDKPSILPQKRNKAQENKLDKEVEEPKTS 840
QY 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLENLNFNMDGTIELYLP 900
Db 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLENLNFNMDGTIELYLP 900
QY 901 SGEVYKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKVPKP 960
Db 901 SGEVYKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKVPKP 960
QY 961 ENSTDNGMLNPEGVNSDPMLEAPFAVDPVQEKLEKFTASYGLGLDSVIFNMDGTI 1020
Db 961 ENSTDNGMLNPEGVNSDPMLEAPFAVDPVQEKLEKFTASYGLGLDSVIFNMDGTI 1020
QY 1021 ELRLPSGEVIKKNLSD 1036
Db 1021 ELRLPSGEVIKKNLSD 1036

RESULT 3

C95136
conserved domain protein Sp1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:gl4972655; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1% Score 32; DB 2; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 52 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 83
|||||

RESULT 4
B95136
conserved domain protein Sp1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B95136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1% Score 32; DB 2; Length 819;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||

RESULT 5
E98004
hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:gl5458683; GSPDB:GN00174
C:Genetics:
A:Gene: phtA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1% Score 32; DB 2; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 78 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 109
|||||

RESULT 6
G95115
conserved hypothetical protein SPI003 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI003
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1% Score 32; DB 2; Length 839;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||

RESULT 7
C97985
hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 853;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSRHHYNGKVPYDA 98
|||||
Db 67 AEQIVIKITDQGYVTSRHHYNGKVPYDA 98
|||||

RESULT 8
D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Lee, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <RUR>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:gl5458682; GSPDB:GN00174
C:Genetics:
A:Gene: bnpA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSRHHYNGKVPYDA 98
|||||
Db 77 AEQIVIKITDQGYVTSRHHYNGKVPYDA 108
|||||

RESULT 9
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; Luedtke, J.; Luedtke, B.; Luedtke, E.; Luedtke, S.; Luedtke, T.; Luedtke, U.; Luedtke, V.; Luedtke, W.; Luedtke, X.; Luedtke, Y.; Luedtke, Z.; Luedtke, AA; Luedtke, AB; Luedtke, AC; Luedtke, AD; Luedtke, AE; Luedtke, AF; Luedtke, AG; Luedtke, AH; Luedtke, AI; Luedtke, AJ; Luedtke, AK; Luedtke, AL; Luedtke, AM; Luedtke, AN; Luedtke, AO; Luedtke, AP; Luedtke, AQ; Luedtke, AR; Luedtke, AS; Luedtke, AT; Luedtke, AU; Luedtke, AV; Luedtke, AW; Luedtke, AX; Luedtke, AY; Luedtke, AZ; Luedtke, BA; Luedtke, BB; Luedtke, BC; Luedtke, BD; Luedtke, BE; Luedtke, BF; Luedtke, BG; Luedtke, BH; Luedtke, BI; Luedtke, BJ; Luedtke, BK; Luedtke, BL; Luedtke, BM; Luedtke, BN; Luedtke, BO; Luedtke, BP; Luedtke, BQ; Luedtke, BR; Luedtke, BS; Luedtke, BT; Luedtke, BU; Luedtke, BV; Luedtke, BW; Luedtke, BX; Luedtke, BY; Luedtke, BZ; Luedtke, CA; Luedtke, CB; Luedtke, CC; Luedtke, CD; Luedtke, CE; Luedtke, CF; Luedtke, CG; Luedtke, CH; Luedtke, CI; Luedtke, CJ; Luedtke, CK; Luedtke, CL; Luedtke, CM; Luedtke, CN; Luedtke, CO; Luedtke, CP; Luedtke, CQ; Luedtke, CR; Luedtke, CS; Luedtke, CT; Luedtke, CU; Luedtke, CV; Luedtke, CW; Luedtke, CX; Luedtke, CY; Luedtke, CZ; Luedtke, DA; Luedtke, DB; Luedtke, DC; Luedtke, DD; Luedtke, DE; Luedtke, DF; Luedtke, DG; Luedtke, DH; Luedtke, DI; Luedtke, DJ; Luedtke, DK; Luedtke, DL; Luedtke, DM; Luedtke, DN; Luedtke, DO; Luedtke, DP; Luedtke, DQ; Luedtke, DR; Luedtke, DS; Luedtke, DT; Luedtke, DU; Luedtke, DV; Luedtke, DW; Luedtke, DX; Luedtke, DY; Luedtke, DZ; Luedtke, EA; Luedtke, EB; Luedtke, EC; Luedtke, ED; Luedtke, EE; Luedtke, EF; Luedtke, EG; Luedtke, EH; Luedtke, EI; Luedtke, EJ; Luedtke, EK; Luedtke, EL; Luedtke, EM; Luedtke, EN; Luedtke, EO; Luedtke, EP; Luedtke, EQ; Luedtke, ER; Luedtke, ES; Luedtke, ET; Luedtke, EU; Luedtke, EV; Luedtke, EW; Luedtke, EX; Luedtke, EY; Luedtke, EZ; Luedtke, FA; Luedtke, FB; Luedtke, FC; Luedtke, FD; Luedtke, FE; Luedtke, FF; Luedtke, FG; Luedtke, FH; Luedtke, FI; Luedtke, FJ; Luedtke, FK; Luedtke, FL; Luedtke, FM; Luedtke, FN; Luedtke, FO; Luedtke, FP; Luedtke, FQ; Luedtke, FR; Luedtke, FS; Luedtke, FT; Luedtke, FU; Luedtke, FV; Luedtke, FW; Luedtke, FX; Luedtke, FY; Luedtke, FZ; Luedtke, GA; Luedtke, GB; Luedtke, GC; Luedtke, GD; Luedtke, GE; Luedtke, GF; Luedtke, GG; Luedtke, GH; Luedtke, GI; Luedtke, GJ; Luedtke, GK; Luedtke, GL; Luedtke, GM; Luedtke, GN; Luedtke, GO; Luedtke, GP; Luedtke, GQ; Luedtke, GR; Luedtke, GS; Luedtke, GT; Luedtke, GU; Luedtke, GV; Luedtke, GW; Luedtke, GX; Luedtke, GY; Luedtke, GZ; Luedtke, HA; Luedtke, HB; Luedtke, HC; Luedtke, HD; Luedtke, HE; Luedtke, HF; Luedtke, HG; Luedtke, HH; Luedtke, HI; Luedtke, HJ; Luedtke, HK; Luedtke, HL; Luedtke, HM; Luedtke, HN; Luedtke, HO; Luedtke, HP; Luedtke, HQ; Luedtke, HR; Luedtke, HS; Luedtke, HT; Luedtke, HU; Luedtke, HV; Luedtke, HW; Luedtke, HX; Luedtke, HY; Luedtke, HZ; Luedtke, IA; Luedtke, IB; Luedtke, IC; Luedtke, ID; Luedtke, IE; Luedtke, IF; Luedtke, IG; Luedtke, IH; Luedtke, II; Luedtke, IJ; Luedtke, IK; Luedtke, IL; Luedtke, IM; Luedtke, IN; Luedtke, IO; Luedtke, IP; Luedtke, IQ; Luedtke, IR; Luedtke, IS; Luedtke, IT; Luedtke, IU; Luedtke, IV; Luedtke, IW; Luedtke, IX; Luedtke, IY; Luedtke, IZ; Luedtke, JA; Luedtke, JB; Luedtke, JC; Luedtke, JD; Luedtke, JE; Luedtke, JF; Luedtke, JG; Luedtke, JH; Luedtke, JI; Luedtke, JJ; Luedtke, JK; Luedtke, JL; Luedtke, JM; Luedtke, JN; Luedtke, JO; Luedtke, JP; Luedtke, JQ; Luedtke, JR; Luedtke, JS; Luedtke, JT; Luedtke, JU; Luedtke, JV; Luedtke, JW; Luedtke, JX; Luedtke, JY; Luedtke, JZ; Luedtke, KA; Luedtke, KB; Luedtke, KC; Luedtke, KD; Luedtke, KE; Luedtke, KF; Luedtke, KG; Luedtke, KH; Luedtke, KI; Luedtke, KL; Luedtke, KM; Luedtke, KN; Luedtke, KO; Luedtke, KP; Luedtke, KQ; Luedtke, KR; Luedtke, KS; Luedtke, KT; Luedtke, KU; Luedtke, KV; Luedtke, KW; Luedtke, KX; Luedtke, KY; Luedtke, KZ; Luedtke, LA; Luedtke, LB; Luedtke, LC; Luedtke, LD; Luedtke, LE; Luedtke, LF; Luedtke, LG; Luedtke, LH; Luedtke, LI; Luedtke, LJ; Luedtke, LK; Luedtke, LL; Luedtke, LM; Luedtke, LN; Luedtke, LO; Luedtke, LP; Luedtke, LQ; Luedtke, LR; Luedtke, LS; Luedtke, LT; Luedtke, LU; Luedtke, LV; Luedtke, LW; Luedtke, LX; Luedtke, LY; Luedtke, LZ; Luedtke, MA; Luedtke, MB; Luedtke, MC; Luedtke, MD; Luedtke, ME; Luedtke, MF; Luedtke, MG; Luedtke, MH; Luedtke, MI; Luedtke, MN; Luedtke, MO; Luedtke, MP; Luedtke, MQ; Luedtke, MR; Luedtke, MS; Luedtke, MT; Luedtke, MU; Luedtke, MV; Luedtke, MW; Luedtke, MX; Luedtke, MY; Luedtke, MZ; Luedtke, NA; Luedtke, NB; Luedtke, NC; Luedtke, ND; Luedtke, NE; Luedtke, NF; Luedtke, NG; Luedtke, NH; Luedtke, NI; Luedtke, NJ; Luedtke, NK; Luedtke, NL; Luedtke, NM; Luedtke, NO; Luedtke, NP; Luedtke, NQ; Luedtke, NR; Luedtke, NS; Luedtke, NT; Luedtke, NU; Luedtke, NV; Luedtke, NW; Luedtke, NX; Luedtke, NY; Luedtke, NZ; Luedtke, OA; Luedtke, OB; Luedtke, OC; Luedtke, OD; Luedtke, OE; Luedtke, OF; Luedtke, OG; Luedtke, OH; Luedtke, OI; Luedtke, OJ; Luedtke, OK; Luedtke, OL; Luedtke, OM; Luedtke, ON; Luedtke, OO; Luedtke, OP; Luedtke, OQ; Luedtke, OR; Luedtke, OS; Luedtke, OT; Luedtke, OU; Luedtke, OV; Luedtke, OW; Luedtke, OX; Luedtke, OY; Luedtke, OZ; Luedtke, PA; Luedtke, PB; Luedtke, PC; Luedtke, PD; Luedtke, PE; Luedtke, PF; Luedtke, PG; Luedtke, PH; Luedtke, PI; Luedtke, PJ; Luedtke, PK; Luedtke, PL; Luedtke, PM; Luedtke, PN; Luedtke, PO; Luedtke, PP; Luedtke, PQ; Luedtke, PR; Luedtke, PS; Luedtke, PT; Luedtke, PU; Luedtke, PV; Luedtke, PW; Luedtke, PX; Luedtke, PY; Luedtke, PZ; Luedtke, QA; Luedtke, QB; Luedtke, QC; Luedtke, QD; Luedtke, QE; Luedtke, QF; Luedtke, QG; Luedtke, QH; Luedtke, QI; Luedtke, QJ; Luedtke, QK; Luedtke, QL; Luedtke, QM; Luedtke, QN; Luedtke, QO; Luedtke, QP; Luedtke, QQ; Luedtke, QR; Luedtke, QS; Luedtke, QT; Luedtke, QU; Luedtke, QV; Luedtke, QW; Luedtke, QX; Luedtke, QY; Luedtke, QZ; Luedtke, RA; Luedtke, RB; Luedtke, RC; Luedtke, RD; Luedtke, RE; Luedtke, RF; Luedtke, RG; Luedtke, RH; Luedtke, RI; Luedtke, RJ; Luedtke, RK; Luedtke, RL; Luedtke, RM; Luedtke, RN; Luedtke, RO; Luedtke, RP; Luedtke, RQ; Luedtke, RR; Luedtke, RS; Luedtke, RT; Luedtke, RU; Luedtke, RV; Luedtke, RW; Luedtke, RX; Luedtke, RY; Luedtke, RZ; Luedtke, SA; Luedtke, SB; Luedtke, SC; Luedtke, SD; Luedtke, SE; Luedtke, SF; Luedtke, SG; Luedtke, SH; Luedtke, SI; Luedtke, SJ; Luedtke, SK; Luedtke, SL; Luedtke, SM; Luedtke, SN; Luedtke, SO; Luedtke, SP; Luedtke, SQ; Luedtke, SR; Luedtke, SS; Luedtke, ST; Luedtke, SU; Luedtke, SV; Luedtke, SW; Luedtke, SX; Luedtke, SY; Luedtke, SZ; Luedtke, TA; Luedtke, TB; Luedtke, TC; Luedtke, TD; Luedtke, TE; Luedtke, TF; Luedtke, TG; Luedtke, TH; Luedtke, TI; Luedtke, TJ; Luedtke, TK; Luedtke, TL; Luedtke, TM; Luedtke, TN; Luedtke, TO; Luedtke, TP; Luedtke, TQ; Luedtke, TR; Luedtke, TS; Luedtke, TT; Luedtke, TU; Luedtke, TV; Luedtke, TW; Luedtke, TX; Luedtke, TY; Luedtke, TZ; Luedtke, UA; Luedtke, UB; Luedtke, UC; Luedtke, UD; Luedtke, UE; Luedtke, UF; Luedtke, UG; Luedtke, UH; Luedtke, UI; Luedtke, UJ; Luedtke, UK; Luedtke, UL; Luedtke, UM; Luedtke, UN; Luedtke, UO; Luedtke, UP; Luedtke, UQ; Luedtke, UR; Luedtke, US; Luedtke, UT; Luedtke, UU; Luedtke, UV; Luedtke, UW; Luedtke, UX; Luedtke, UY; Luedtke, UZ; Luedtke, VA; Luedtke, VB; Luedtke, VC; Luedtke, VD; Luedtke, VE; Luedtke, VF; Luedtke, VG; Luedtke, VH; Luedtke, VI; Luedtke, VJ; Luedtke, VK; Luedtke, VL; Luedtke, VM; Luedtke, VN; Luedtke, VO; Luedtke, VP; Luedtke, VQ; Luedtke, VR; Luedtke, VS; Luedtke, VT; Luedtke, VU; Luedtke, VV; Luedtke, VW; Luedtke, VX; Luedtke, VY; Luedtke, VZ; Luedtke, WA; Luedtke, WB; Luedtke, WC; Luedtke, WD; Luedtke, WE; Luedtke, WF; Luedtke, WG; Luedtke, WH; Luedtke, WI; Luedtke, WJ; Luedtke, WK; Luedtke, WL; Luedtke, WM; Luedtke, WN; Luedtke, WO; Luedtke, WP; Luedtke, WQ; Luedtke, WR; Luedtke, WS; Luedtke, WT; Luedtke, WU; Luedtke, WV; Luedtke, WX; Luedtke, WY; Luedtke, WZ; Luedtke, XA; Luedtke, XB; Luedtke, XC; Luedtke, XD; Luedtke, XE; Luedtke, XF; Luedtke, XG; Luedtke, XH; Luedtke, XI; Luedtke, XJ; Luedtke, XK; Luedtke, XL; Luedtke, XM; Luedtke, XN; Luedtke, XO; Luedtke, XP; Luedtke, XQ; Luedtke, XR; Luedtke, XS; Luedtke, XT; Luedtke, XU; Luedtke, XV; Luedtke, XW; Luedtke, XX; Luedtke, XY; Luedtke, XZ; Luedtke, YA; Luedtke, YB; Luedtke, YC; Luedtke, YD; Luedtke, YE; Luedtke, YF; Luedtke, YG; Luedtke, YH; Luedtke, YI; Luedtke, YJ; Luedtke, YK; Luedtke, YL; Luedtke, YM; Luedtke, YN; Luedtke, YO; Luedtke, YP; Luedtke, YQ; Luedtke, YR; Luedtke, YS; Luedtke, YT; Luedtke, YU; Luedtke, YV; Luedtke, YW; Luedtke, YX; Luedtke, YY; Luedtke, YZ; Luedtke, ZA; Luedtke, ZB; Luedtke, ZC; Luedtke, ZD; Luedtke, ZE; Luedtke, ZF; Luedtke, ZG; Luedtke, ZH; Luedtke, ZI; Luedtke, ZJ; Luedtke, ZK; Luedtke, ZL; Luedtke, ZM; Luedtke, ZN; Luedtke, ZO; Luedtke, ZP; Luedtke, ZQ; Luedtke, ZR; Luedtke, ZS; Luedtke, ZT; Luedtke, ZU; Luedtke, ZV; Luedtke, ZW; Luedtke, ZX; Luedtke, ZY; Luedtke, ZZ

Query Match 1.5%; Score 16; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSRHHYNGKVPYDA 88
|||||
Db 75 KITDQGYVTSRHHYNGKVPYDA 90
|||||

RESULT 10
E97985
hypothetical protein spr0909 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: E97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Lee, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-46 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:gl5458516; GSPDB:GN00174
C:Genetics:
A:Gene: spr0909

Query Match 0.9%; Score 9; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 IVPHGHHYH 210
|||||
Db 21 IVPHGHHYH 29
|||||

RESULT 11
F97985
hypothetical protein phtE-truncation [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Lee, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:gl5458517; GSPDB:GN00174
C:Genetics:
A:Gene: phtE-truncation

Query Match 0.9%; Score 9; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 391 AYIVRHGDH 399
|||||
Db 30 AYIVRHGDH 38
|||||

RESULT 12
C69541
conserved hypothetical protein AF2331 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69541
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kickness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: C69541
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

C:Date: 1-92 <KLE>
A:Cross-references: GB:AE000943; GB:AE000782; NID:g2689266; PIDN:AA88923.1; PID:g264818

Query Match 0.8%; Score 8; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 843 VEKEKLS 850
Db 78 VEKEKLS 85
|||||

RESULT 13
A61291
ferredoxin [2Fe-2S] - parsley
C:Species: Petroselinum crispum (parsley)
C:Date: 27-Jun-1994 #sequence_revision 19-May-1995 #text_change 07-May-1999
A:Accession: A61291
R:Nakano, T.; Hase, T.; Matsubara, H.
J. Biochem. 90, 1725-1730, 1981
A:Title: The complete amino acid sequence of parsley (Petroselinum sativum) ferredoxin.
A:Reference number: A61291; MUID:82142254; PMID:7334006
A:Accession: A61291
A:Molecule type: protein
A:Residues: 1-96 <NAK>
A:Note: the source is designated as Petroselinum sativum
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:24-78/Domain: ferredoxin [2Fe-2S] homology <FER>
F:39,44,47,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 5 VKLITPDG 12
|||||

RESULT 14
FEED
ferredoxin [2Fe-2S] - European elder (tentative sequence)
C:Species: Sambucus nigra (European elder)
C:Date: 30-Apr-1980 #sequence_revision 30-Apr-1980 #text_change 31-Mar-2000
A:Accession: A00233
R:Takurai, I.A.H.; Boulter, D.
Phytochemistry 18, 1481-1484, 1979
A:Title: The amino acid sequence of ferredoxin from Sambucus nigra.
A:Reference number: A00233
A:Accession: A00233
A:Molecule type: protein
A:Residues: 1-97 <TAK>
A:Note: the amidation states of residues 57, 58, 60, 61, 68, and 88 were identified by H
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:1-97/Product: ferredoxin [2Fe-2S] #status experimental <MAT>
F:24-78/Domain: ferredoxin [2Fe-2S] homology <FER>
F:39,44,47,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 5 VKLITPDG 12
|||||

RESULT 15
FEYB6
ferredoxin [2Fe-2S] [validated] - Synechocystis sp.
C:Species: Synechocystis sp.

C:Date: 13-Jun-1983 #sequence_revision 13-Nov-1998 #text_change 15-Sep-2000
A:Accession: S76345; A56811; A00247
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, O.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76345
A:Molecule type: DNA
A:Residues: 1-97 <KAN>
A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BA010197.1; PID:g100
A:Experimental source: strain PCC 6803
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
R:Bottin, H.; Lagoutte, B.
Biochim. Biophys. Acta 1101, 48-56, 1992
A:Title: Ferredoxin and flavodoxin from the cyanobacterium Synechocystis sp PCC 6803.
A:Reference number: A56811; MUID:92338182; PMID:1633177
A:Accession: A56811
A:Molecule type: protein
A:Residues: 2-97 <BOT>
A:Experimental source: strain PCC 6714
A:Note: sequence extracted from NCBI backbone (NCBIP:109680)
R:Hase, T.; Inoue, K.; Matsubara, H.; Williams, M.M.; Rogers, L.J.
J. Biochem. 92, 1357-1362, 1982
A:Title: Amino acid sequence of Synechocystis 6714 ferredoxin: a unique structural fe
A:Reference number: A00247; MUID:83108768; PMID:6818221
A:Accession: A00247
A:Molecule type: protein
A:Residues: 2-14, N, 16-97 <HAS>
A:Experimental source: strain PCC 6714
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
submitted to the Brookhaven Protein Data Bank, September 1995
A:Reference number: A65450; PDB:IDOX
A:Contents: annotation; conformation and disulfide bond assignments by (1)H- and (15)
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
submitted to the Brookhaven Protein Data Bank, September 1995
A:Reference number: A65451; PDB:IDOX
A:Contents: annotation; conformation and disulfide bond assignments by (1)H- and (15)
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
Biochemistry 34, 14462-14473, 1995
A:Title: (1)H and (15)N NMR sequential assignment, secondary structure, and tertiary
A:Reference number: A58608; MUID:96062510; PMID:7578051
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:2-97/Product: ferredoxin [2Fe-2S] #status experimental <MAT>
F:25-79/Domain: ferredoxin [2Fe-2S] homology <FER>
F:19-86/Disulfide bonds: #status predicted
F:40,45,48,78/Binding site: 2Fe-2S cluster (Cys) (covalent) #status experimental

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 6 VKLITPDG 13
|||||

RESULT 16
S39689
vwd protein - Bacillus subtilis
N:Alternate names: hypothetical protein ipa-34d
C:Species: Bacillus subtilis
C:Date: 07-Oct-1994 #sequence_rev#sion 26-May-1995 #text_change 15-Oct-1999
A:Accession: S39689; G70052
R:Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionesc
A.: Rapoport, G.; Danchin, A.
Mol. Microbiol. 10, 371-384, 1993
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region
A:Reference number: S39655; MUID:95020537; PMID:7934828
A:Accession: S39689

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-127 <GLA>
 A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51590.1; PID:g413958
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Enrich, S.D.; Emerson, P.T.; Encian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A95580; MUID:98044033; PMID:9384377
 A:Accession: G70052
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-127 <KUN>
 A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15847.1; PID:ell86320;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ywcd
 C:Keywords: transmembrane protein

Query Match 0.8%; Score 8; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 LOSLLKEL 275
 |||||
 DB 59 LOSLLKEL 66

RESULT 17
 FERZ
 ferredoxin [2Fe-2S] I precursor - rice
 C:Species: Oryza sativa (rice)
 C:Date: 31-Mar-1989 #sequence_revision 28-May-1999 #text_change 16-Jun-2000
 A:Accession: T03738; S03730; JTO223
 R:Ohnori, K.; Doyama, N.; Ida, S.
 Plant Physiol. 111, 348, 1996
 A:Title: Molecular cloning of a rice leaf ferredoxin cDNA.
 A:Reference number: Z15043
 A:Accession: T03738
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-139 <OHM>
 A:Cross-references: EMBL:D30763; PIDN:BA06436.1
 A:Experimental source: subsp. japonica, cv. Kinmaze
 R:Kamo, M.; Kotani, N.; Tsugita, A.; He, Y.K.; Nozu, Y.
 Protein Seq. Data Anal. 2, 289-293, 1989
 A:Title: Amino acid sequences of ferredoxins from rice cultivars, japonica and indica.
 A:Reference number: S03730; MUID:89367259; PMID:2771933
 A:Accession: S03730
 A:Molecule type: protein
 A:Residues: 44-139 <KAM>
 A>Note: sequences from cultivars japonica and indica are identical
 C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
 C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
 F:44-139/Product: ferredoxin [2Fe-2S] I #status experimental <MAT>
 F:67-121/Domain: ferredoxin [2Fe-2S] homology <FER>
 F:82,87,90,120/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

QY 641 VKLITPDG 648
 |||||
 DB 48 VKLITPDG 55

RESULT 18
 A96919
 PTS system, mannitol-specific IIA domain (Ntr-type) (gene MltF) [imported] - Clostrid
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: A96919
 R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: A96919
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78140.1; PID:g15022985; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0156
 C:Superfamily: phosphotransferase system mannitol-specific enzyme II factor III; phos

Query Match 0.8%; Score 8; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
 |||||
 DB 58 NGVAIPHG 55

RESULT 19
 AD3475
 protein-Npi-phosphohistidine-sugar phosphotransferase (EC 2.7.1.65) [imported] - Bruc
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
 C:Accession: AD3475
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
 ; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melit*
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AD3475
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL52967.1; PID:g17983818; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI1786
 A:Map position: 1
 C:Superfamily: phosphotransferase system enzyme II; phosphotransferase system mannitol
 C:Keywords: phosphotransferase

Query Match 0.8%; Score 8; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
 |||||
 DB 60 NGVAIPHG 67

RESULT 20
 A69349
 conserved hypothetical transmembrane protein AF0793 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

A:Accession: JQ0270
 A:Molecule type: DNA
 A:Residues: 1-273 <SHI>
 A:Experimental source: cv. Nihonbare
 A:Author: H. Shimada, H. Whittier, R.; Ishibashi, T.; Sakamoto, M.; Mori, M.; Ko
 M.
 Mol. Gen. Genet. 217, 185-194, 1989
 A:Title: The complete sequence of the rice (Oryza sativa) chloroplast genome: intermo
 of the cereals.
 A:Reference number: S05080; MUID:89364698; PMID:2770692
 A:Accession: S05150
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-273 <HIR>
 A:Experimental source: cv. Nihonbare
 A:Note: This sequence was submitted to EMBL, July 1989
 A:Author: R. R. Moon, E.; Wu, R.
 Gene 70, 1-12, 1988
 A:Title: Organization and nucleotide sequence of genes at both junctions between the
 A:Reference number: JAO092; MUID:89198901; PMID:3248862
 A:Accession: JAO093
 A:Molecule type: DNA
 A:Residues: 1-68, 'E', 70-199, 'K', 201-229, 'MGAVKGRKPLVEKNPQ', 247-273 <MOO>
 A:Genetics:
 A:Gene: rpl2
 A:Genome: chloroplast
 A:Start codon: ACG
 A:Introns: 131/1
 A:Superfamily: Escherichia coli ribosomal protein L2
 A:Keywords: chloroplast; protein biosynthesis; ribosome
 Query Match 0.8%; Score 8; DB 1; Length 273;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1022 LRLPSGEV 1029
 |||||
 Db 172 LRLPSGEV 179
 RESULT 23
 R52M2
 A:ribosomal protein L2 - maize chloroplast
 A:Species: chloroplast Zea mays (maize)
 A:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 21-Jul-2000
 A:Accession: S10500; S17874; S58638; S58595
 A:Author: K.; Giese, K.; Larrinua, I.M.; McLaughlin, W.E.; Subramanian, A.R.
 Nucleic Acids Res. 18, 4244, 1990
 A:Title: Nucleotide sequence and map positions of the duplicated gene for maize (Zea
 A:Reference number: S10500; MUID:90332419; PMID:2377464
 A:Accession: S10500
 A:Molecule type: DNA
 A:Residues: 1-273 <KAV>
 A:Cross-references: EMBL:X53066
 A:Author: Hoch, B.; Maier, R.M.; Appel, K.; Igloi, G.L.; Koessel, H.
 Nature 353, 178-180, 1991
 A:Title: Editing of a chloroplast mRNA by creation of an initiation codon.
 A:Reference number: S17874; MUID:91367263; PMID:1653905
 A:Accession: S17874
 A:Molecule type: mRNA
 A:Residues: 1-150 <HOC>
 A:Cross-references: EMBL:X62070
 A:Author: Maier, R.M.; Neckermann, K.; Igloi, G.L.; Koessel, H.
 J. Mol. Biol. 251, 614-628, 1995
 A:Title: Complete sequence of the maize chloroplast genome: gene content, hotspots of
 A:Reference number: S58531; MUID:95395841; PMID:7666415
 A:Accession: S58639
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-273 <MAI>
 A:Cross-references: EMBL:X86563; NID:9902200; PIDN:CAA60329.1; PID:g2673858
 A:Genetics: GEN1
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, Dec 11 1995

A:Accession: S58595
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <MAN>
A:Cross-references: EMBL:X86563; NID:9022200; PIDN:CAA50329.1; PID:g2673858
A:Genetics: GEN2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics: <GEN1>
A:Map position: IR(A)
A:Genome: chloroplast
A:Start codon: ACG
A:Introns: 130/3
A:Note: the start codon AUG results from mRNA editing of ACG
C:Genetics: <GEN2>
A:Map position: IR(B)
A:Genome: chloroplast
A:Introns: 130/3
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome; RNA editing

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 172 LRLPSGEV 179

RESULT 24
T11810
ribosomal protein L2 - Norway spruce chloroplast
C:Species: chloroplast Picea abies (Norway spruce)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T11810
R:Kluemper, S.; Kanka, S.; Riesner, D.; Etscheid, M.
Submitted to the EMBL Data Library, March 1997
A:Description: Characterisation of a Norway spruce chloroplast DNA clone: Complete nucleotide sequence
A:Reference number: Z17349
A:Accession: T11810
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-275 <KLU>
A:Cross-references: EMBL:U92462; NID:g2959581; PID:g2959586
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Introns: 133/1
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 0.8%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 174 LRLPSGEV 181

RESULT 25
T07361
ribosomal protein L2 - Chlorella vulgaris chloroplast
C:Species: chloroplast Chlorella vulgaris
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07361
R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, S.; Tsudzuki, J.; Nakas
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlorella vulgaris
A:Reference number: Z15985; MUID:97303241; PMID:9159184
A:Accession: T07361
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-275 <WAK>
A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAA58009.1; PID:g2224525
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; ribosome

Query Match 0.8%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 174 LRLPSGEV 181

RESULT 26
T07531
ribosomal protein L2 - Japanese black pine chloroplast (fragment)
C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C:Accession: T07531
R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugitara, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
A:Reference number: Z16030; MUID:95024047; PMID:7937893
A:Accession: T07531
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <WAK>
A:Cross-references: EMBL:DI7510; NID:g529643; PIDN:BAA23474.1; PID:g2626947
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Note: Intron positions not resolved (incomplete sequence)
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; ribosome

Query Match 0.8%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 175 LRLPSGEV 182

RESULT 27
AF2987
ABC transporter, membrane spanning protein Atu3503 [imported] - Agrobacterium tumefac
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF2987
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2987
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <KUR>
A:Cross-references: GB:AB008689; PIDN:AAL44316.1; PID:gl7741906; GSPDB:GN00187
C:Genetics:
A:Gene: Atu3503
A:Map position: linear chromosome
C:Superfamily: maltose transport protein malG

A:Accession: S58595
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <MAN>
A:Cross-references: EMBL:X86563; NID:9022200; PIDN:CAA50329.1; PID:g2673858
A:Genetics: GEN2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics: <GEN1>
A:Map position: IR(A)
A:Genome: chloroplast
A:Start codon: ACG
A:Introns: 130/3
A:Note: the start codon AUG results from mRNA editing of ACG
C:Genetics: <GEN2>
A:Map position: IR(B)
A:Genome: chloroplast
A:Introns: 130/3
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome; RNA editing

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 172 LRLPSGEV 179

RESULT 24
T11810
ribosomal protein L2 - Norway spruce chloroplast
C:Species: chloroplast Picea abies (Norway spruce)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T11810
R:Kluemper, S.; Kanka, S.; Riesner, D.; Etscheid, M.
Submitted to the EMBL Data Library, March 1997
A:Description: Characterisation of a Norway spruce chloroplast DNA clone: Complete nucleotide sequence
A:Reference number: Z17349
A:Accession: T11810
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-275 <KLU>
A:Cross-references: EMBL:U92462; NID:g2959581; PID:g2959586
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Introns: 133/1
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 0.8%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
|||||
Db 174 LRLPSGEV 181

RESULT 25
T07361
ribosomal protein L2 - Chlorella vulgaris chloroplast
C:Species: chloroplast Chlorella vulgaris
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07361
R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, S.; Tsudzuki, J.; Nakas
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlorella vulgaris
A:Reference number: Z15985; MUID:97303241; PMID:9159184
A:Accession: T07361
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

Query Match          0.8%; Score 8; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 LDPALPEA 988
    |||||
Db 166 LDPALPEA 173

RESULT 28
B98296
sulfate ABC transporter, permease protein VC0539 [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: B98296
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: B98296
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89892.1; PID:g15159839; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L2652
A:Map position: linear chromosome
C:Superfamily: maltose transport protein malG

Query Match          0.8%; Score 8; DB 2; Length:278;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 LDPALPEA 988
    |||||
Db 167 LDPALPEA 174

RESULT 29
AI3633
hypothetical protein BMEII0994 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AI3633
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AI3633
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54236.1; PID:g17985208; GSPDB:GN00191
C:Genetics:
A:Gene: BMEII0994
A:Map position: II

Query Match          0.8%; Score 8; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 PNGVAIPH 309
    |||||
Db 258 PNGVAIPH 265

RESULT 30
H59094
hypothetical protein pX01-32 - Bacillus anthracis virulence plasmid pX01
C:Species: Bacillus anthracis

```

```

C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: H59094
R:Oknaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483; PMID:10515943
A:Accession: H59094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <OXI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32336.1; PID:g4894248
A:Experimental source: strain Sterne
C:Genetics:
A:Gene: pX01-32
A:Genome: plasmid
C:Superfamily: Bacillus anthracis virulence plasmid pX01 hypothetical protein pX01-32

Query Match          0.8%; Score 8; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 587 EGNKVYTG 594
    |||||
Db 154 EGNKVYTG 161

RESULT 31
G83148
DNA polymerase III, delta subunit PA3989 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83148
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <STO>
A:Cross-references: GB:AE004816; GB:AE004091; NID:g9950168; PIDN:AAG07376.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: hola; PA3989

Query Match          0.8%; Score 8; DB 2; Length 345;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1020 IELRLPSG 1027
    |||||
Db 79 IELRLPSG 86

RESULT 32
AD2087
gas vesicle protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2087
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2087
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE73950.1; PID:g17131342; GSPDB:GN00179

```

A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: gvpN

Query Match 0.8% Score 8; DB 2; Length 410;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 LSALEEKI 328
DB 145 LSALEEKI 152

RESULT 33

F81329
serine proteinase (proteinase DO) (EC 3.4.21.-) Cj1228c [imported] - Campylobacter jejuni

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81329
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, W.P.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, N.; et al. Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: F81329

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-472 <PAR>

A:Cross-references: GB:AL119077; GB:AL111168; NID:g6968444; PIDN:CAB73482.1; PID:g696866

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: htrA; Cj1228c

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase, serine proteinase

Query Match 0.8% Score 8; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 EVVSSLSGS 356
DB 98 EVVSSLSGS 105

RESULT 34

B83735

cassette chromosome recombinase B1 BH0682 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B83735

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y. Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its relationship to other Bacillus species

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B83735

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-522 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA804401.1; GSPDB:GN000004

C:Genetics:

A:Gene: BH0682

Query Match 0.8% Score 8; DB 2; Length 522;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 KLDEKVEE 836
DB 203 KLDEKVEE 210

RESULT 35

B97331

membrane associated methyl-accepting chemotaxis protein with HAMP domain [imported]

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97331

R:Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97331

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81437.1; PID:g15026604; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3511

Query Match 0.8% Score 8; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 NVVNLKKN 606
DB 259 NVVNLKKN 266

RESULT 36

S61157

hypothetical protein YDR362c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D9476.13

C:Species: Saccharomyces cerevisiae

C>Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002

C:Accession: S61157

R:Du, Z.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of S. cerevisiae cosmid 9476.

A:Reference number: S61148

A:Accession: S61157

A:Molecule type: DNA

A:Residues: 1-672 <DUZ>

A:Cross-references: EMBL:U28372; NID:g849170; PID:g849183; GSPDB:GN000004; MIPS:YDR362

C:Genetics:

A:Gene: SGD:YDR362c

A:Cross-references: SGD:S0002770

A:Map position: 4R

C:Superfamily: Saccharomyces cerevisiae hypothetical protein YDR362c

Query Match 0.8% Score 8; DB 2; Length 672;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 YIFNPKDI 384
DB 492 YIFNPKDI 499

RESULT 37

S48261

hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0821

C:Species: Saccharomyces cerevisiae

C>Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 13-Apr-2002

C:Accession: S48261; S45962; S41800; S44676

R:Mannhaupt, G.; Stucka, R.; Ehle, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48261

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-753 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55599.1; PID:g476052
 R:Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45927
 A:Accession: S45962
 A:Molecule type: DNA
 A:Residues: 1-753 <FE2>
 A:Cross-references: EMBL:Z35963; NID:g536366; PIDN:CAA85047.1; PID:g536367; MIPS:YBR094
 R:Decker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
 submitted to the EMBL Data Library, December 1992
 A:Reference number: S31313
 A:Accession: S41800
 A:Molecule type: DNA
 A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
 A:Cross-references: EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID:g4088
 C:Genetics:
 A:Cross-references: SGD:S0000298
 A:Map position: 2R

Query Match 0.8%; Score 8; DB 2; Length 753;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 KAHLAGKN 233
 DB 50 KAHLAGKN 57
 |||||

RESULT 38
 T18280
 hypothetical protein D2 - slime mold (Dictyostellium discoideum)
 C:Species: Dictyostellium discoideum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 A:Accession: T18280
 R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh
 Genetics 148, 1117-1125, 1998
 A:Title: Dictyostellium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1
 A:Reference number: Z14684; MUID:98198836; PMID:9539429
 A:Accession: T18280
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-763 <RIE>
 A:Cross-references: EMBL:U00796; NID:g2702254; PID:g2702255; PIDN:AAC18631.1

Query Match 0.8%; Score 8; DB 2; Length 763;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 NMLVKLIT 645
 DB 632 NMLVKLIT 639
 |||||

RESULT 39
 T40674
 protein transport protein sec23 homolog - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T40674
 R:Lynne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: Z21944
 A:Accession: T40674
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-765 <LYN>
 A:Cross-references: EMBL:AL035263; PIDN:CAA22877.1; GSPDB:GN00067; SPDB:SPBC776.04
 A:Experimental source: strain 972h-; cosmid c776
 C:Genetics:
 A:Gene: SPDB:SPBC776.04
 A:Map position: 2
 A:Introns: 11/1; 58/2; 79/1
 C:Superfamily: SEC23 protein

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55599.1; PID:g476052
 R:Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45927
 A:Accession: S45962
 A:Molecule type: DNA
 A:Residues: 1-753 <FE2>
 A:Cross-references: EMBL:Z35963; NID:g536366; PIDN:CAA85047.1; PID:g536367; MIPS:YBR094
 R:Decker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
 submitted to the EMBL Data Library, December 1992
 A:Reference number: S31313
 A:Accession: S41800
 A:Molecule type: DNA
 A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
 A:Cross-references: EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID:g4088
 C:Genetics:
 A:Cross-references: SGD:S0000298
 A:Map position: 2R

Query Match 0.8%; Score 8; DB 2; Length 753;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 KAHLAGKN 233
 DB 50 KAHLAGKN 57
 |||||

RESULT 38
 T18280
 hypothetical protein D2 - slime mold (Dictyostellium discoideum)
 C:Species: Dictyostellium discoideum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 A:Accession: T18280
 R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh
 Genetics 148, 1117-1125, 1998
 A:Title: Dictyostellium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1
 A:Reference number: Z14684; MUID:98198836; PMID:9539429
 A:Accession: T18280
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-763 <RIE>
 A:Cross-references: EMBL:U00796; NID:g2702254; PID:g2702255; PIDN:AAC18631.1

Query Match 0.8%; Score 8; DB 2; Length 763;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 NMLVKLIT 645
 DB 632 NMLVKLIT 639
 |||||

RESULT 39
 T40674
 protein transport protein sec23 homolog - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T40674
 R:Lynne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: Z21944
 A:Accession: T40674
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-765 <LYN>
 A:Cross-references: EMBL:AL035263; PIDN:CAA22877.1; GSPDB:GN00067; SPDB:SPBC776.04
 A:Experimental source: strain 972h-; cosmid c776
 C:Genetics:
 A:Gene: SPDB:SPBC776.04
 A:Map position: 2
 A:Introns: 11/1; 58/2; 79/1
 C:Superfamily: SEC23 protein

Query Match 0.8%; Score 8; DB 2; Length 765;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 AVIVSLSL 20
 DB 144 AVIVSLSL 151
 |||||

RESULT 40
 G96494
 protein F7F22.4 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G96494
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: G96494
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-813 <STO>
 A:Cross-references: GB:AE005173; NID:g6691183; PIDN:AAF24521.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F7F22.4
 A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 813;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIK 479
 DB 432 DLTEQIK 439
 |||||

RESULT 41
 C75313
 probable proteinase - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: C75313
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: C75313
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-840 <WHI>
 A:Cross-references: GB:AF002047; GB:AE000513; NID:g6459915; PIDN:AAF11680.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2130
 A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 840;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 DGLVFDPA 295
 |||||

```

Db      332 DGLVDPDA 339
RESULT 42
G86393
protein T24Pl3.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86393
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cready, I.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-984 <STO>
A:Cross-references: GB:AE005172; NID:g295721; PIDN:AAF87027.1; GSPDB:GN00141
C:Genetics:
A:Gene: T24Pl3.6
A:Map position: 1

Query Match      0.8%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 SLSSNPSS 363
      |||||
Db      843 SLSSNPSS 850

RESULT 43
S72635
exo-poly-alpha-galacturonosidase (EC 3.2.1.82) precursor - Thermoanaerobacterium thermos
C:Species: Thermoanaerobacterium thermosulfurigenes
C:Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 15-Oct-1999
C:Accession: S72635; S72621
R:Matuschek, M.; Salm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A:Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes
A:Reference number: S72635
A:Accession: S72635
A:Molecule type: DNA
A:Residues: 1-1148 <MATU>
A:Cross-references: EMBL:U50951; NID:g1542972; PIDN:AAB08040.1; PID:g1255234
R:Experimental source: strain EM1
R:Matuschek, M.; Salm, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A:Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 tha
A:Reference number: S72621; MUID:97033555; PMID:8879252
A:Accession: S72621
A:Molecule type: DNA
A:Residues: 729-1148 <MAW>
A:Cross-references: EMBL:U50951
A:Experimental source: EM1
C:Genetics:
A:Gene: palA
C:Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1148/Product: exo-poly-alpha-galacturonosidase #status predicted <MAT>
F:959-1022/Domain: S-layer repeat homology <SLR1>
F:1028-1081/Domain: S-layer repeat homology <SLR2>
F:1092-1144/Domain: S-layer repeat homology <SLR3>

Query Match      0.8%; Score 8; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 LKNSTFNN 611
      |||||
Db      598 LKNSTFNN 605

RESULT 44
TL4265
golgin-245 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: TL4265
R:Gay, D.; Beller, B.M.; Zhao, H.; Bhandoola, A.; Tomayko, M.M.; Murali, R.; Marks, M.
submitted to the EMBL Data Library, February 1998
A:Description: Olp-1, the murine golgin-245/p230 homologue, is upregulated during oli
A:Reference number: Z17947
A:Accession: TL4265
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1218 <GAY>
A:Cross-references: EMBL:AF051357; NID:g2952521; PID:g2952522; PIDN:AAC05573.1
A:Experimental source: strain BALB/C; brain
C:Genetics:
A:Gene: olp-1

Query Match      0.8%; Score 8; DB 2; Length 1218;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 SAVIVSL 19
      |||||
Db      114 SAVIVSL 121

RESULT 45
T45070
protein kinase homolog R31240_1 [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45070
R:Lamerdin, J.; McCready, P.; Stilwagen, S.; Ramirez, M.; Carrano, A.
submitted to the EMBL Data Library, November 1996
A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb re
A:Reference number: Z22906
A:Accession: T45070
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1237 <LAM>
A:Cross-references: EMBL:AD000092; PIDN:AAB51171.1
A:Experimental source: cell line 5HL2-B; fibroblast
C:Genetics:
A:Map position: 19p13.2
A:Note: Intron positions not resolved (incomplete sequence)

Query Match      0.8%; Score 8; DB 2; Length 1237;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 SLSSNPSS 363
      |||||
Db      595 SLSSNPSS 602

Search completed: May 13, 2003, 14:01:41
Job time : 35 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:56:37 ; Search time 16 Seconds
(without alignments)
2693.369 Million cell updates/sec

Title: US-09-471-255-2
Perfect score: 1039
Sequence: 1 MKFSKKYIAGSAVIVSL.....IELPLPGEVKKNLSDFFIA 1039

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 8

Total number of hits satisfying chosen parameters: 15

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.8	92	1 YN31_ARCFU	O27953 archaeoglob
2	8	0.8	96	1 FER1_ORYSA	P11051 oryza sativ
3	8	0.8	96	1 FER_SYN34	P27320 synechocyst
4	8	0.8	96	1 FER_SYN4	P00243 synechocyst
5	8	0.8	97	1 FER_SAMN1	P00226 sambucus ni
6	8	0.8	127	1 YWCD_BACSU	P39602 bacillus su
7	8	0.8	233	1 RL1_THEMA	P29393 thermotoga
8	8	0.8	273	1 RK2_HORVU	P41086 hordeum vul
9	8	0.8	273	1 RK2_MAIZE	P17788 zea mays (m
10	8	0.8	273	1 RK2_ORYSA	P17351 oryza sativ
11	8	0.8	275	1 RK2_CHLYU	P56367 chlorella v
12	8	0.8	275	1 RK2_PICAB	O62954 picea abies
13	8	0.8	276	1 RK2_PINTH	O62940 pinus thunb
14	8	0.8	335	1 GVPN_ANAFL	P55150 anabaena fl
15	8	0.8	753	1 YBU4_YEAST	P38234 saccharomyc

ALIGNMENTS

RESULT 1

ID	YN31_ARCFU	STANDARD;	PRT;	92 AA.
AC	O27953;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Hypothetical protein AF2331.			
CN	AF2331.			
OS	Archaeoglobus fulgidus.			
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;			
OC	Archaeoglobaceae; Archaeoglobus.			
OX	NCBI_TaxID=2234;			

SEQUENCE FROM N.A.

STRAIN=VC-16 / DSM 4304 / ATCC 49558;

MEDLINE=98049343; Pubmed-9389475;

Kienk H.-P.; Clayton R.A.; Tomb J.-F.; White O.; Nelson K.E.; Kechum K.A.; Dodson R.J.; Winn M.; Hickey E.K.; Peterson J.D.; Richardson D.L.; Kerlavage A.R.; Graham D.E.; Kyriades N.C.; Fleischmann R.D.; Quackenbush J.; Lee N.H.; Sutton G.G.; Gill S.; Kirkness E.F.; Dougherty B.A.; McKenney K.; Adams M.D.; Loftus B.; Peterson S.; Reich C.I.; McNeil L.K.; Badger J.H.; Glodek A.; Zhou L.; Overbeek R.; Gocayne J.D.; Weidman J.F.; McDonald L.; Utterback T.; Cotton M.D.; Spriggs T.; Artiach P.; Kaine B.P.; Sykes S.M.; Sadow P.W.; D'Andrea K.P.; Bowman C.; Fujii C.; Garland S.A.; Mason T.M.; Olsen G.J.; Fraser C.M.; Smith H.O.; Woese C.R.; Venter J.C.;

"The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus."

Nature 390:364-370(1997)

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE000943; AAB88923.1; -

TIGR; AF2331; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 92 AA; 10521 MW; CCC444D2796013CC CRC64;

Query Match 0.88; Score 8; DB 1; Length 92;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 843 VEKEKLSK 850

DB 78 VEKEKLSK 85

RESULT 2

ID	FER1_ORYSA	STANDARD;	PRT;	96 AA.
AC	P11051;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Ferredoxin I.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehrhartoideae; Oryzaceae; Oryza.			
OX	NCBI_TaxID=4530;			
RN	[1]			
RP	SEQUENCE.			
RC	STRAIN=cv. Japonica, and cv. Indica;			
RX	MEDLINE=89367259; Pubmed=2771933;			
RA	Kamo M.; Kotani N.; Tsugita A.; He Y.K.; Nozu Y.;			
RT	"Amino acid sequences of ferredoxins from rice cultivars, Japonica			
RT	and Indica."			
RL	Protein Seq. Data Anal. 2:289-293(1989).			
CC	-!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER			
CC	ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.			
CC	-!- COFACTOR: BINDS 1 2FE-2S CLUSTER.			
CC	-!- SUBCELLULAR LOCATION: Chloroplast.			
DR	PIR; J02023; FERZ			
DR	PIR; S03730; S03730.			
DR	HSSP; P00221; 1A70.			
DR	InterPro; IPR000564; 2Fe2S_ferredoxin.			
DR	InterPro; IPR001041; Ferredoxin.			
DR	Pfam; PF00111; fer2; 1.			
DR	PRINTS; PR00159; 2FE2SFRDXIN.			

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: D85607; BAA24020.1; -
DR EMBL: U38802; AAB72025.1; -
DR EMBL: D64000; BAA10197.1; -
DR PDB: 1DOX; 08-MAR-96.
DR PDB: IDOY; 08-MAR-96.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDOXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur; 3D-structure; Complete proteome.
FT INIT_MET 0
FT METAL 39 39 IRON-SULFUR (2FE-2S).
FT METAL 44 44 IRON-SULFUR (2FE-2S).
FT METAL 47 47 IRON-SULFUR (2FE-2S).
FT METAL 77 77 IRON-SULFUR (2FE-2S).
FT DISULFID 18 85 PROBABLE.
SQ SEQUENCE 96 AA; 10232 MW; 73866DD68E369451 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 5 VKLITPDG 12

RESULT 3
FER_SYNY3 STANDARD; PRT; 96 AA.
AC P27320;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferredoxin I
GN PEIF OR FED OR SLL0011.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamoto H., Suzuki T.;
RT "Cloning, characterization and transcriptional studies of ferredoxin
RT genes from the mesophilic cyanobacterium Synechocystis sp. PCC 6803
RT and the thermophilic cyanobacterium Synechococcus vulcanus.";
RL Physiol. Plantarum 101:199-203(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Cassier-Chauvat C., Poncelet M., Viljoing S., Chauvat F.;
RL Submitted (APR-1996) to the EMBL/GenBank/DDJB databases.
RN [4]
RP SEQUENCE.
RX MEDLINE=92338182; PubMed=1633177;
RA Bottin H., Lagoutte B.;
RT "Ferredoxin and flavodoxin from the cyanobacterium Synechocystis sp
RT PCC 6803.";
RL Biochim. Biophys. Acta 1101:48-56(1992).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE=96062510; PubMed=7578051;
RA Lelong C., Setif P., Bottin H., Andre F., Neumann J.-M.;
RT "1H and 15N NMR sequential assignment, secondary structure, and
RT tertiary fold of [2Fe-2S] ferredoxin from Synechocystis sp. PCC
RT 6803.";
RL Biochemistry 34:14462-14473(1995).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC -!- MISCELLANEOUS: THE MIDPOINT REDOX POTENTIAL FOR THIS PROTEIN IS
CC -412 MV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

Query Match 0.8%; Score 8; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 5 VKLITPDG 12

RESULT 4
FER_SYNY4 STANDARD; PRT; 96 AA.
AC P00243;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1988 (Rel. 09, Last annotation update)
DE Ferredoxin.
OS Synechocystis sp. (strain PCC 6714).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1147;
RN [1]
RP SEQUENCE.
RX MEDLINE=83108768; PubMed=6818221;
RA Hase T., Inoue K., Matsubara H., Williams M.M., Rogers L.J.;
RT "Amino acid sequence of Synechocystis 6714 ferredoxin: a unique
RT structural feature of unicellular blue-green algal ferredoxin.";
RL J. Biochem. 92:1357-1362(1982).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC HSP: P27320; 1DOX.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDOXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur.
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 44 44 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 77 77 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 96 AA; 10259 MW; 5E86781964135BBB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648


```

Db      5 VKLITPDG 12
|||||
RESULT 5
FER_SAMMI
ID FER_SAMMI STANDARD; PRT; 97 AA.
AC P00226;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Ferredoxin.
OS Sambucus nigra (European elder).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Dipsacales; Adoxaceae; Sambucus.
OX NCBI_TaxID=4202;
RN [1]
RP SEQUENCE.
RA Takurai I.A.H., Boulter D.;
RT "The amino acid sequence of ferredoxin from Sambucus nigra.";
RL Phytochemistry 18:1481-1484(1979).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COPACITOR: BINDS 1 2FE-2S CLUSTER.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
DR PIR; A00233; FEED.
DR HSSP; P00221; 1A70.
DR InterPro; IPR000564; 2Fe2S_ferredoxin.
DR InterPro; IPR001041; Ferredoxin.
DR Pfam; PF00111; fer2; 1.
DR PRINTS; PF00159; 2FE2SPRDOXIN
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur; Chloroplast.
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 44 44 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 77 77 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 97 AA; 10616 MW; D19C34A411C68AA2 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
|||||
Db      5 VKLITPDG 12
|||||

RESULT 6
YWCD_BACSU
ID YWCD_BACSU STANDARD; PRT; 127 AA.
AC P39602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ywcd.
GN YWCD OR IPA-34D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Repoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caidwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serot S.J., Serrif P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein F., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE GTRA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAAS1590.1; -.
DR EMBL; Z99123; CAB15847.1; -.
DR PIR; S39689; S39689.
DR Subtilist; BG10580; ywcd.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
SQ SEQUENCE 127 AA; 14349 MW; 1A51311C0A0FA05F CRC64;

Query Match 0.8%; Score 8; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 LQSLKEL 275
|||||
Db      59 LQSLKEL 66
|||||

RESULT 7
RLA_THEME
ID RLI_THEME STANDARD; PRT; 233 AA.
AC P29393;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L1.

```

```

GN RPLA OR TM0455.
OC Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=93054590; PubMed=1429627;
RA Dennis P.P., Liao D.;
RT "The organization and expression of essential transcription
RT translation component genes in the extremely thermophilic eubacterium
RT Thermotoga maritima.";
RL J. Biol. Chem. 267:22787-22797(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.W.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC LOCATED IN THE NEIGHBORHOOD OF THE SITE WHERE ELONGATION FACTOR TU
CC IS BOUND TO THE RIBOSOME (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LIP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11839; CAA77860.1;
DR HSP; P04257; IRL2;
DR InterPro; IPR002171; Ribosomal_L2;
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRfams; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
DR Ribosomal protein; Chloroplast.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 273 AA; 30045 MW; 4C204966ACB7DBE9 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179

RESULT 9
RK2_MAIZE
ID RK2_MAIZE STANDARD; PRT; 273 AA.
AC PI7788;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.
OS Zea mays (Maize).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, PR9CMSSR37;
RX MEDLINE=9032419; PubMed=2377464;
RA Kavousi M., Giese K., Larrinua I.M., Subramanian A.R.;
RT "Nucleotide sequence and map positions of the duplicated gene for
RT maize (Zea mays) chloroplast ribosomal protein L2.";
RL Nucleic Acids Res. 18:4244-4244(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9535841; PubMed=7666415;
RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;
RT "Complete sequence of the maize chloroplast genome: gene content,
RT hotspots of divergence and fine tuning of genetic information by
RT transcript editing.";
RL J. Mol. Biol. 251:614-628(1995).

```

```

RN  RNA EDITING OF INITIATOR CODON.
RX  MEDLINE=91367263; PubMed=1653905;
RA  Hoch B., Maier R.M., Appel K., Igloi G.L., Koessel H.;
RT  "Editing of a chloroplast mRNA by creation of an initiation codon.";
CC  -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC  -1- CAUTION: THE INITIATOR METHIONINE IS CREATED BY RNA EDITING.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X53066; CA37241.1;
DR  EMBL; X86563; CA60329.1; ALT_SEQ.
DR  EMBL; X86563; CA60371.1;
DR  EMBL; X62070; CA43983.1; ALT_SEQ.
DR  PIR; S10500; R52N2.
DR  PIR; S17874; S17874.
DR  HSP; P04257; IRL2.
DR  MaizeDB; 66413;
DR  InterPro; IPR002171; Ribosomal_L2.
DR  Pfam; PF00181; Ribosomal_L2; 1.
DR  TIGRFAMs; TIGR01171; rplB_bact; 1.
DR  PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW  Ribosomal protein; Chloroplast; RNA editing.
RW  Ribosomal protein; Chloroplast; RNA editing.
SQ  SEQUENCE 273 AA; 30065 MW; BA65197231EA3CA0 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179

RESULT 10
RK2_ORYSA STANDARD; PRT; 273 AA.
AC P17351; P92327;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2-A AND RPL2-B.
OS Oryza sativa (Rice).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89196901; PubMed=3240862;
RA Moon E., Wu R.;
RT "Organization and nucleotide sequence of genes at both junctions
RT between the two inverted repeats and the large single-copy region in
RT the rice chloroplast genome.";
RL Gene 70:1-12(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX MEDLINE=89364698; PubMed=2770692;
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,
RA Kanno A., Nishizawa Y., Hirai A., Shinozaki K., Sugura M.;
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:
RT intermolecular recombination between distinct trna genes accounts for
```

```

a major plastid DNA inversion during the evolution of the cereals.";
RL Mol. Gen. Genet. 217:185-194(1989).
CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M22826; NOT_ANNOTATED_CDS.
DR  EMBL; L40578; AAD15254.1; ALT_SEQ.
DR  EMBL; X15901; CA43924.1;
DR  EMBL; X15901; CA43924.1; ALT_SEQ.
DR  PIR; JQ0270; R5R22.
DR  HSP; P04257; IRL2.
DR  InterPro; IPR002171; Ribosomal_L2.
DR  Pfam; PF00181; Ribosomal_L2; 1.
DR  TIGRFAMs; TIGR01171; rplB_bact; 1.
DR  PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW  Ribosomal protein; Chloroplast.
RW  Ribosomal protein; Chloroplast.
FT CONFLICT 69 69 Q -> E (IN REF. 1).
FT CONFLICT 200 200 Q -> K (IN REF. 1).
FT CONFLICT 230 246 HGGEGCKADIGKKPTT -> GAVKGGKPLPEKNPQP
FT CONFLICT (IN REF. 1).
SQ SEQUENCE 273 AA; 29942 MW; AA153062AA8729B CRC64;

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179

RESULT 11
RK2_CHLVU STANDARD; PRT; 275 AA.
AC P56367;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.
OS Chlorella vulgaris.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Sugura M.;
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorella vulgaris: the existence of genes possibly
RT involved in chloroplast division.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC  -----
```

```

DR EMBL; AB001684; BAA58009.1; -
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 275 AA; 30093 MW; 369FF92300441165 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 174 LRLPSGEV 181

RESULT 12
RK2_PICAB
ID RK2_PICAB STANDARD; PRT; 275 AA.
AC O62954;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.
OS Picea abies (Norway spruce) (Picea excelsa).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3329;
RN [1]
RP Kluemper S., Kanka S., Riesner D., Etscheid M.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U92462; AAC95500.1; -
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; FALSE_NEG.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 275 AA; 30180 MW; 1DFDE72CE798C378 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 174 LRLPSGEV 181

RESULT 13
RK2_PINTH
ID RK2_PINTH STANDARD; PRT; 276 AA.
AC O62940;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.

```

```

OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17510; BAA23474.1; -
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; FALSE_NEG.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 276 AA; 30278 MW; 7F783A55998FB146 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 175 LRLPSGEV 182

RESULT 14
GVPN_ANAFL
ID GVPN_ANAFL STANDARD; PRT; 335 AA.
AC P55150;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gas vesicle protein gvpN.
GN GVPN.
OS Anabaena flos-aquae.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=1166;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CCAP 1403/13f;
RX MEDLINE=97217585; PubMed=9063646;
RA Kinsman R., Hayes P.K.;
RT "Genes encoding proteins homologous to halobacterial Gvps N, J, K, F
RT & L are located downstream of gvpC in the cyanobacterium Anabaena
RT flos-aquae";
RL DNA Seq. 7:97-106(1997).
CC -!- SIMILARITY: BELONGS TO THE CBQ/NITQ/NORQ/GVPN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U17109; AAA58711.1; -

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:57:43 ; Search time 48 Seconds
(without alignments)
4460.062 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKKYTAGSAVIVLSL.....IELRLPGEVIKKNLSDFIA 1039

Scoring table: OLICO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 98

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	834	80.3	1039	16 Q9ANV1	Q9anv1 streptococc
2	32	3.1	802	16 Q97QM8	Q97qm8 streptococc
3	32	3.1	816	2 Q9AHY9	Q9ahy9 streptococc
4	32	3.1	819	2 Q9ANY3	Q9any3 streptococc
5	32	3.1	819	16 Q97QM9	Q97qm9 streptococc
6	32	3.1	839	16 Q9ANY2	Q9any2 streptococc
7	32	3.1	844	2 Q9AG74	Q9ag74 streptococc
8	16	1.5	289	2 Q9AE21	Q9ae21 streptococc
9	16	1.5	822	2 Q9ZHG7	Q9zhg7 streptococc
10	16	1.5	825	2 Q93GT5	Q93gt5 streptococc
11	16	1.5	825	16 Q99XV4	Q99xv4 streptococc
12	9	0.9	198	5 Q8SSL6	Q8ssl6 encephalito
13	9	0.9	336	10 Q9LIM0	Q9lim0 arabidopsis
14	8	0.8	40	8 Q8WIJ2	Q8wij2 bostrychia
15	8	0.8	46	8 Q8WIJ0	Q8wij0 bostrychia
16	8	0.8	47	8 Q8WIJ8	Q8wij8 bostrychia

17	8	0.8	8	Q8WII9	bostrychia
18	8	0.8	47	Q8WII6	bostrychia
19	8	0.8	50	Q8WII6	bostrychia
20	8	0.8	50	Q8WII4	bostrychia
21	8	0.8	50	Q8WII3	bostrychia
22	8	0.8	50	Q8WII1	bostrychia
23	8	0.8	50	Q8WII8	bostrychia
24	8	0.8	50	Q8W929	bostrychia
25	8	0.8	50	Q8W823	bostrychia
26	8	0.8	50	Q8W822	bostrychia
27	8	0.8	50	Q8W8W8	bostrychia
28	8	0.8	50	Q8W8W7	bostrychia
29	8	0.8	131	5 Q95VU3	Q95vu3 homarus ame
30	8	0.8	138	10 Q8S9N2	Q8s9n2 oryza sativ
31	8	0.8	139	2 Q65991	Q65991 clostridium
32	8	0.8	139	10 Q40683	Q40683 oryza sativ
33	8	0.8	139	10 Q22382	Q22382 oryza sativ
34	8	0.8	144	10 Q9ZITS2	Q9zits2 capsicum an
35	8	0.8	144	10 Q93XJ9	Q93xj9 solanum tub
36	8	0.8	146	16 Q97MN7	Q97mn7 clostridium
37	8	0.8	146	16 Q8RCS1	Q8rcs1 thermoanaer
38	8	0.8	154	16 Q8YEU1	Q8yeu1 bruceella me
39	8	0.8	182	8 Q8GFJ0	Q8gfj0 ginkgo bilo
40	8	0.8	185	8 Q8GFJ9	Q8gfj9 illicium pa
41	8	0.8	190	8 Q8GFJ2	Q8gfj2 dioscorea b
42	8	0.8	190	8 Q8GFJ6	Q8gfj6 saururus ce
43	8	0.8	199	8 Q8GFJ5	Q8gfj5 calycanthus
44	8	0.8	200	8 Q8GFJ6	Q8gfj6 cabomba car
45	8	0.8	201	8 Q8GFJ8	Q8gfj8 acorus cala
46	8	0.8	201	8 Q8GFJ7	Q8gfj7 asarum cana
47	8	0.8	201	8 Q8GFJ4	Q8gfj4 ceratophyll
48	8	0.8	201	8 Q8GFJ1	Q8gfj1 drimys wint
49	8	0.8	201	8 Q8GFJ8	Q8gfj8 lactoris fe
50	8	0.8	201	8 Q8GFJ7	Q8gfj7 liriodendro
51	8	0.8	201	8 Q8MSQ5	Q8msq5 zamia furfu
52	8	0.8	201	8 Q8MSP3	Q8msp3 nymphaea od
53	8	0.8	201	8 Q8GE23	Q8ge23 ambarella t
54	8	0.8	207	17 Q29465	Q29465 archaeoglob
55	8	0.8	252	15 Q9IUI8	Q9iui8 human immun
56	8	0.8	252	15 Q9IUI3	Q9iui3 human immun
57	8	0.8	274	16 Q9RXB3	Q9rbx3 streptococ
58	8	0.8	278	16 Q8YBA6	Q8yba6 bruceella me
59	8	0.8	278	16 Q8UA71	Q8ua71 agrobacteri
60	8	0.8	312	10 Q94HD4	Q94hd4 oryza sativ
61	8	0.8	314	12 Q91MW4	Q91mw4 lumpy skin
62	8	0.8	327	2 Q9X303	Q9x303 bacillus an
63	8	0.8	345	16 Q9HX31	Q9hx31 pseudomonas
64	8	0.8	372	5 Q9ULI8	Q9uli8 drosophila
65	8	0.8	410	16 Q8YU70	Q8yut0 anabaena sp
66	8	0.8	412	17 Q96XE9	Q96xe9 sulfolobus
67	8	0.8	472	2 Q46120	Q46120 campylobact
68	8	0.8	472	16 Q9PN69	Q9pn69 campylobact
69	8	0.8	514	2 Q9XSR9	Q9xsr9 streptomyce
70	8	0.8	522	16 Q9KEF5	Q9kef5 bacillus ha
71	8	0.8	543	4 Q9BY87	Q9by87 homo sapien
72	8	0.8	564	2 Q9R9P9	Q9r9p9 rhizobium m
73	8	0.8	573	16 Q97DG6	Q97dg6 clostridium
74	8	0.8	661	11 Q9JKS3	Q9jks3 mus musculu
75	8	0.8	672	3 Q66339	Q66339 saccharomyc
76	8	0.8	681	4 Q9UPI0	Q9upio homo sapien
77	8	0.8	695	10 Q8WSP0	Q8wsp0 oryza sativ
78	8	0.8	723	11 Q9VWH2	Q9vwh2 mus musculu
79	8	0.8	723	11 Q9JKS4	Q9jks4 mus musculu
80	8	0.8	743	5 Q23859	Q23859 dictyosteli
81	8	0.8	763	5 Q23874	Q23874 dictyosteli
82	8	0.8	765	3 Q94672	Q94672 schizosacch
83	8	0.8	778	5 Q8TIF5	Q8tif5 dictyosteli
84	8	0.8	791	16 Q9CLO1	Q9cl01 pasteurella
85	8	0.8	793	10 Q9ASU6	Q9asu6 arabidopsis
86	8	0.8	813	10 Q9ASH6	Q9ash6 arabidopsis
87	8	0.8	840	15 Q9RSJ4	Q9rsj4 deirococcus
88	8	0.8	984	10 Q9LQY4	Q9lqy4 arabidopsis
89	8	0.8	1100	5 Q9VUE7	Q9vue7 drosophila

90 8 0.8 1148 2 Q60045 Q60045 thermoanaer
91 8 0.8 1210 11 Q91VW5 Q91VW5 mus musculus
92 8 0.8 1237 4 Q00114 Q00114 homo sapien
93 8 0.8 1308 4 Q60307 Q60307 homo sapien
94 8 0.8 1353 3 Q8X1F2 Q8X1F2 venturia in
95 8 0.8 1566 11 Q9R1L5 Q9R1L5 mus musculus
96 8 0.8 1583 4 Q9V2H9 Q9V2H9 homo sapien
97 8 0.8 2238 11 Q70365 Q70365 mus musculus
98 8 0.8 2986 12 Q91BL4 Q91BL4 perina nuda

ALIGNMENTS

RESULT 1
Q9ANY1 PRELIMINARY; PRT: 1039 AA.
AC Q9ANY1
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004)
GN PTE OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.,
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis."
RL Infect. Immun. 69:949-958(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Mayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., C.L.,
RA Holtzapple E., Khouri H., Wolf A.M., Otterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae".
RL Science 293:498-506(2001)
DR EMBL: AF318956; AA06761.1;
DR EMBL: AE007403; AA075121.1;
DR TIGR: SP1004;
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 POTENTIAL
SQ SEQUENCE 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match 80.3%; Score 834; DB 16; Length 1039;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1034; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKFSKYIAGSAVIVSLSCAVALNHRQENKDNRRVYVDSOSSKSENLTDPQVS 60
DB 1 MKFSKYIAGSAVIVSLSCAVALNHRQENKDNRRVYVDSOSSKSENLTDPQVS 60
QY 61 QKEGIAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDALFSELLMKDPNQLKADIVN 120
DB 61 QKEGIAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDALFSELLMKDPNQLKADIVN 120
QY 121 EVKGGYIIKVDGKYVYLKDAHADNVRTKDEINRQKQEHVKDNKNSNVAVARSQRY 180
DB 121 EVKGGYIIKVDGKYVYLKDAHADNVRTKDEINRQKQEHVKDNKNSNVAVARSQRY 180

DB 121 EVKGGYIIKVDGKYVYLKDAHADNVRTKDEINRQKQEHVKDNKNSNVAVARSQRY 180
QY 181 TTNDGYVFNPAIDIEDTGNAYIVPHGGHYHYIPKSDLSASASAAKAHLAKGNQPSOLS 240
DB 181 TTNDGYVFNPAIDIEDTGNAYIVPHGGHYHYIPKSDLSASASAAKAHLAKGNQPSOLS 240
QY 241 YSSTASDNNTQSAVAGSTSKPANKSENQLSKELYPSPSAQRYSESGLVDFPAKIIIS 300
DB 241 YSSTASDNNTQSAVAGSTSKPANKSENQLSKELYPSPSAQRYSESGLVDFPAKIIIS 300
QY 301 TPNGVAIPHGHDHYHFIPIYSKLSALEEKIARMVPIISGTGSTVSTNAKNEVVSSLSLSSN 360
DB 301 TPNGVAIPHGHDHYHFIPIYSKLSALEEKIARMVPIISGTGSTVSTNAKNEVVSSLSLSSN 360
QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLA 420
DB 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLA 420
QY 421 TPSPSLP-INPGTSHKHEEDGYGFDANRIIAEDSESGFVMSHGDHNYFFKKDLTEEQIKA 480
DB 421 TPSPSLP-INPGTSHKHEEDGYGFDANRIIAEDSESGFVMSHGDHNYFFKKDLTEEQIKA 480
QY 481 AQKHEEVKTSHGSLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVVN 540
DB 481 AQKHEEVKTSHGSLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVVN 540
QY 541 KEKNAIIPHGHDHHPIDHFKPVGIGHSHSNVLELFPKEEGVAKKGNKYVTEELTNV 600
DB 541 KEKNAIIPHGHDHHPIDHFKPVGIGHSHSNVLELFPKEEGVAKKGNKYVTEELTNV 600
QY 601 VNLKNSTFNQNFNLANGQKRVSPFPPELEKLGINMLVKLITPDGKVLKESGKVFG 660
DB 601 VNLKNSTFNQNFNLANGQKRVSPFPPELEKLGINMLVKLITPDGKVLKESGKVFG 660
QY 661 EGVGNIANFELDQPLPGQTKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHA 720
DB 661 EGVGNIANFELDQPLPGQTKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHA 720
QY 721 GDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPKPLNOGTTTACNK 780
DB 721 GDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPKPLNOGTTTACNK 780
QY 781 IPIVTMANAYLDNQSTYIIVEPILKEKQTDKPSILPQFKRKAQENSKLDEKVEEPKTS 840
DB 781 IPIVTMANAYLDNQSTYIIVEPILKEKQTDKPSILPQFKRKAQENSKLDEKVEEPKTS 840
QY 841 ERVEKEKLSGTNSNSTSLTEVPVDPVQKVAFAESYGNKLENVLFNMDGTIELYLP 900
DB 841 ERVEKEKLSGTNSNSTSLTEVPVDPVQKVAFAESYGNKLENVLFNMDGTIELYLP 900
QY 901 SGEVIKKMAADFTGEAPOGNGENKPSENGKSTGTVENQPTENPADSLPEAPNEKPVKP 960
DB 901 SGEVIKKMAADFTGEAPOGNGENKPSENGKSTGTVENQPTENPADSLPEAPNEKPVKP 960
QY 961 ENSTDNGMLNPEGVSGDPMLEAPADVPDVEKLEKFTASYGLGLDSVIFNMDGTI 1020
DB 961 ENSTDNGMLNPEGVSGDPMLEAPADVPDVEKLEKFTASYGLGLDSVIFNMDGTI 1020
QY 1021 EURLPSGEVIKKNLSD 1036
DB 1021 EURLPSGEVIKKNLSD 1036
RESULT 2
Q97OM8 PRELIMINARY; PRT: 802 AA.
AC Q97OM8
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.


```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL; AE007418; AAK75284.1;
DR TIGR; SP1175;
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1833 CRC64;

Query Match 3.1%; Score 32; DB 16; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
DB 52 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 83

RESULT 3
Q9AHT9 ID Q9AHT9 PRELIMINARY; PRT; 816 AA.
AC Q9AHT9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wismemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RA "Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection Against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1;
DR TIGR; SP1175;
KW Complete proteome.
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 3.1%; Score 32; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
DB 66 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 97

RESULT 4
Q9ANY3 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.
AC Q9ANY3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Pneumococcal histidine triad protein B precursor (Fragment).
GN PHTB.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dornitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RA "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
DR EMBL; AF318954; AAK06759.1;
DR TIGR; SP1174;
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 819
FT SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 3.1%; Score 32; DB 2; Length 819;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
DB 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98

RESULT 5
Q97QM9 ID Q97QM9 PRELIMINARY; PRT; 819 AA.
AC Q97QM9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conserved domain protein.
GN SP1174.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL; AE007418; AAK75283.1;
DR TIGR; SP1174;
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 3.1%; Score 32; DB 16; Length 819;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
DB 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98

```

2

```

KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 1.5%; Score 16; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDIYH 88
DB 75 KITDQGYVTSHGDIYH 90
|||||

RESULT 10
Q93GT5 PRELIMINARY; PRT; 825 AA.
AC Q93GT5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Terao Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB073859; BAB1774.1; -
SQ SEQUENCE 825 AA; 92623 MW; D64ECC199181DFFB CRC64;

Query Match 1.5%; Score 16; DB 2; Length 825;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDIYH 88
DB 75 KITDQGYVTSHGDIYH 90
|||||

RESULT 11
Q93XV4 PRELIMINARY; PRT; 825 AA.
AC Q93XV4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein SPY2006.
GN SPY2006.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP370 / ATCC 700294 / SEROTYPE M1;
RA MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren O., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4638-4653(2001).
DR EMBL: AE006623; AAK34688.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF50969E6D50F4 CRC64;

Query Match 1.5%; Score 16; DB 16; Length 825;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;

KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 1.5%; Score 16; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDIYH 88
DB 75 KITDQGYVTSHGDIYH 90
|||||

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDIYH 88
DB 75 KITDQGYVTSHGDIYH 90
|||||

RESULT 12
Q8SS16 PRELIMINARY; PRT; 198 AA.
AC Q8SS16;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 60S ribosomal protein L13A (L16).
GN EC004.1380.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thonarat F.,
RA Prenter G., Barbe V., Peyretallade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL: AL590444; CAP25327.1; -
SQ SEQUENCE 198 AA; 22679 MW; 267BC81D1C78FE2B CRC64;

Query Match 0.9%; Score 9; DB 5; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 KDLKRIEE 521
DB 171 KDLKRIEE 179
|||||

RESULT 13
Q9LIM0 PRELIMINARY; PRT; 336 AA.
AC Q9LIM0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Emb|CAB09999.1.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
```

RL DNA Res. 7:217-221(2000).
 DR EMBL; AF382906; AAL67234.1; .
 SQ SEQUENCE 336 AA: 37575 MW; 4BCAC3731EB66F43 CRC64;
 Query Match 0.8%; Score 9; DB 10; Length 336;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 804 LEKENOTDK 812
 DB 133 LEKENOTDK 141
 RESULT 14
 Q8WIJ2 PRELIMINARY; PRT; 40 AA.
 AC Q8WIJ2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3880.QLD;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382902; AAL67226.1; .
 DR InterPro; IPR000894; RuBisCO_small.
 DR Pfam; PF00101; RuBisCO_small; 1.
 DR ProDom; PD000290; RuBisCO_small; 1.
 KW Chloroplast.
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA: 4723 MW; 96013F36C758F617 CRC64;
 Query Match 0.8%; Score 8; DB 8; Length 40;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 472 DLTEEQIK 479
 DB 13 DLTEEQIK 20
 RESULT 15
 Q8WIJ0 PRELIMINARY; PRT; 46 AA.
 AC Q8WIJ0;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P2962.SG;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."

RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382906; AAL67234.1; .
 DR InterPro; IPR000894; RuBisCO_small.
 DR Pfam; PF00101; RuBisCO_small; 1.
 DR ProDom; PD000290; RuBisCO_small; 1.
 KW Chloroplast.
 FT NON_TER 46 46
 SQ SEQUENCE 46 AA: 5345 MW; 6E8EDA7D7D489216 CRC64;
 Query Match 0.8%; Score 8; DB 8; Length 46;
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 472 DLTEEQIK 479
 DB 13 DLTEEQIK 20
 RESULT 16
 Q8WIJ8 PRELIMINARY; PRT; 47 AA.
 AC Q8WIJ8;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.E349.PA;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382893; AAL67208.1; .
 DR InterPro; IPR000894; RuBisCO_small.
 DR Pfam; PF00101; RuBisCO_small; 1.
 DR ProDom; PD000290; RuBisCO_small; 1.
 KW Chloroplast.
 FT NON_TER 47 47
 SQ SEQUENCE 47 AA: 5575 MW; 65F19EFC0529892 CRC64;
 Query Match 0.8%; Score 8; DB 8; Length 47;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 472 DLTEEQIK 479
 DB 13 DLTEEQIK 20
 RESULT 17
 Q8WIJ9 PRELIMINARY; PRT; 47 AA.
 AC Q8WIJ9;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=P4097.MW;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
  (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
  mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382907; AAL67236.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 47 AA; 5508 MW; 65F8ED47D7D4892 CRC64;

  Query Match          0.8%; Score 8; DB 8; Length 47;
  Best Local Similarity 100.0%; Pred. No. 9.3;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 18
Q8WIJ6
ID Q8WIJ6 PRELIMINARY; PRT; 47 AA.
AC Q8WIJ6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3054.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
  (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
  mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382909; AAL67240.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 47 AA; 5573 MW; 8B3C91CFEC7643AA CRC64;

  Query Match          0.8%; Score 8; DB 8; Length 47;
  Best Local Similarity 100.0%; Pred. No. 9.3;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 19
Q8WIJ6
ID Q8WIJ6 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.

```

```

OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3042.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
  (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
  mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382897; AAL67216.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 5985 MW; 1394B315F18E2680 CRC64;

  Query Match          0.8%; Score 8; DB 8; Length 50;
  Best Local Similarity 100.0%; Pred. No. 9.8;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 20
Q8WIJ4
ID Q8WIJ4 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3041.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
  (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
  mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382898; AAL67218.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6001 MW; 655E3D5D5F18E2690 CRC64;

  Query Match          0.8%; Score 8; DB 8; Length 50;
  Best Local Similarity 100.0%; Pred. No. 9.8;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 21
Q8WIJ3
ID Q8WIJ3 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ3;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3065.BZ;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382899; AAL67220.1; -
 DR InterPro; IPR000894; Rubisco_small.
 DR Pfam; PF00101; Rubisco_small; 1.
 DR ProDom; PD000290; Rubisco_small; 1.
 DR Chloroplast.
 DR NON_TER
 FT SEQUENCE 50 AA; 6004 MW; 13854315F18E2680 CRC64;
 SQ

Query Match 0.8%; Score 8; DB 8; Length 50;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
 Db 13 DLTEEQIK 20

RESULT 22
 Q8WJ1 PRELIMINARY; PRT; 50 AA.
 AC Q8WJ1;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P2961.SG;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382905; AAL67232.1; -
 DR InterPro; IPR000894; Rubisco_small.
 DR Pfam; PF00101; Rubisco_small; 1.
 DR ProDom; PD000290; Rubisco_small; 1.
 DR Chloroplast.
 DR NON_TER
 FT SEQUENCE 50 AA; 5937 MW; 2CBA4315FE8ED47D CRC64;
 SQ

Query Match 0.8%; Score 8; DB 8; Length 50;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
 Db 13 DLTEEQIK 20

RESULT 23

Q8WJ18 PRELIMINARY; PRT; 50 AA.
 AC Q8WJ18;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3065.BZ;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382908; AAL67238.1; -
 DR InterPro; IPR000894; Rubisco_small.
 DR Pfam; PF00101; Rubisco_small; 1.
 DR ProDom; PD000290; Rubisco_small; 1.
 DR Chloroplast.
 DR NON_TER
 FT SEQUENCE 50 AA; 5948 MW; 23E5337B222F610C CRC64;
 SQ

Query Match 0.8%; Score 8; DB 8; Length 50;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
 Db 13 DLTEEQIK 20

RESULT 24
 Q8W929 PRELIMINARY; PRT; 50 AA.
 AC Q8W929;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Ribulose-1,5-bisphosphate small subunit (Fragment).
 GN RBCS.
 OS Bostrychia calliptera.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
 OC Bostrychia.
 OX NCBI_TaxID=161377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3190.CO, C3194.CO, C3252.MX, P4122.FL, C3400.MX, AND C3191.CO;
 RA Zuccarello G.C., West J.A.;
 RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
 RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
 RT mitochondrial and plastid DNA markers."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF382890; AAL67202.1; -
 DR EMBL; AF382891; AAL67204.1; -
 DR EMBL; AF382892; AAL67206.1; -
 DR EMBL; AF382894; AAL67210.1; -
 DR EMBL; AF382895; AAL67212.1; -
 DR EMBL; AF382896; AAL67214.1; -
 DR InterPro; IPR000894; Rubisco_small.
 DR Pfam; PF00101; Rubisco_small; 1.
 DR ProDom; PD000290; Rubisco_small; 1.
 DR Chloroplast.
 DR NON_TER
 FT SEQUENCE 50 AA; 6004 MW; 13854315F19EFCE0 CRC64;
 SQ

```

DR EMBL; AF382915; AAL67252.1; -.
DR EMBL; AF382916; AAL67254.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
DR KW Chloroplast.
FT NON_TER 50
SQ SEQUENCE 50 AA; 6060 MW; 23E5337B3EFC5FEC CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 27
Q8W8W8 PRELIMINARY; PRT; 50 AA.
ID Q8W8W8
AC Q8W8W8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2812.QLD, AND P2817.QLD;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382900; AAL67222.1; -.
DR EMBL; AF382901; AAL67224.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
DR KW Chloroplast.
FT NON_TER 50
SQ SEQUENCE 50 AA; 6009 MW; 2CBA430B0D0C447D CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 28
Q8W8W7 PRELIMINARY; PRT; 50 AA.
ID Q8W8W7
AC Q8W8W7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-P128.NT, AND P3752.NT;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF382903; AAL67228.1;
DR EMBL: AF382904; AAL67230.1;
DR InterPro: IPR000894; Rubisco_small.
DR Pfam: PF00101; Rubisco_small; 1.
DR Prodom: PD000290; Rubisco_small; 1.
KW Chloroplast.
FT NON_TER 50 50
SQ SEQUENCE 50 AA; 5995 MW; 2CBA4315FCE3447D CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIK 479
Db 13 DLTEQIK 20

RESULT 29
Q95VU3 PRELIMINARY; PRT: 131 AA.
AC Q95VU3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Olfactory enriched transcript 09.47 (Fragment).
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
OX NCBI_TaxID=6706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OLFACTORY ORGAN;
RA McClintock T.S., Schweder D., Hollins B.;
RT "Olfactory enriched transcripts are cell type specific markers in the
RT lobster olfactory organ."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF405437; AAL04104.1;
FT NON_TER 1 1
FT NON_TER 131 131
SQ SEQUENCE 131 AA; 15093 MW; D9F37C14BDAL2A9F CRC64;

Query Match 0.8%; Score 8; DB 5; Length 131;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 LKDADIVN 120
Db 111 LKDADIVN 118

RESULT 30
Q8S9N2 PRELIMINARY; PRT: 138 AA.
AC Q8S9N2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Anti-disease protein 1.
GN ADII.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;

RP SEQUENCE FROM N.A.
RA Rihe P., Qunhong Y., Aisheng X., Xian L., Huiguo F.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY072818; AAL77198.1;
SQ SEQUENCE 138 AA; 14864 MW; 0A184EE3BCEB94A CRC64;

Query Match 0.8%; Score 8; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 47 VKLITPDG 54

RESULT 31
O65991 PRELIMINARY; PRT: 139 AA.
AC O65991;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MTLF.
GN MTLF.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 792;
RX MEDLINE=21097245; Pubmed=11160802;
RA Behrens S., Mitchell W.J., Bahl H.;
RT "Molecular analysis of the mannitol operon of Clostridium
RT acetobutylicum encoding a phosphotransferase system and a putative
RT PTS-modulated regulator."
RL Microbiology 147:75-86(2001).
DR EMBL: U53868; AAC12850.1;
DR HSSP: P00550; 1A3A.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR Prodom: PD001689; PTS_EIIA_2; 1.
DR PROSITE: PS00372; PTS_EIIA_2; 1.
SQ SEQUENCE 139 AA; 15676 MW; 09471CAC67903B11 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
Db 58 NGVAIPHG 65

RESULT 32
Q40683 PRELIMINARY; PRT: 139 AA.
AC Q40683;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Ferredoxin.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KINMAZE; TISSUE=LEAF;
RA Ohmori K., Doyama N., Ida S.;
RT "Molecular cloning of a rice leaf ferredoxin cDNA.";


```

RL Plant Physiol. 111:348-348(1996).
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: D30763; BAA06436.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDOXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 139 AA; 14951 MW; 301F478A1265E93C CRC64;

Query Match          0.8%; Score 8; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 48 VKLITPDG 55

RESULT 33
Q22382
ID O22382 PRELIMINARY; PRT; 139 AA.
AC O22382:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Ferredoxin.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MILYANG 23; TISSUE=IMMATURE SEED;
RA Lee M.C., Kim C.S., Yi B.Y., Eun M.Y.;
RT "Molecular cloning and characterization of ferredoxin gene from
RT rice.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: AF010320; AAB65699.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDOXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 139 AA; 14945 MW; 4236DF708E02C0D2 CRC64;

Query Match          0.8%; Score 8; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 48 VKLITPDG 55

RESULT 34
Q9ZTS2
ID Q9ZTS2 PRELIMINARY; PRT; 144 AA.
AC Q9ZTS2:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Ferredoxin-like protein.
GN APL.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECW;
RA Lin H.J., Lee B.H., Lee Y.S., Pai C.H., Lin N.S., Feng T.Y.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: AF039562; AAD02175.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDOXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 144 AA; 15178 MW; E9763663281D58EB CRC64;

Query Match          0.8%; Score 8; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 52 VKLITPDG 59

RESULT 35
Q93XJ9
ID Q93XJ9 PRELIMINARY; PRT; 144 AA.
AC Q93XJ9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ferredoxin I.
GN FED1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Holtgreife S., Bader K.P., Horton P., Scheibe R., von Schaewen A.,
RA Backhausen J.E.;
RT "Alterations of leaf ferredoxin (Fd I) contents in transgenic potato
RT (Solanum tuberosum L.) plants changes electron distribution, and
RT influences light acclimation of transgenic plants.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ307031; CAC38395.1; -.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; UNKNOWN_1.
KW Iron-sulfur.
FT CHAIN 48 144 FERREDOXIN I.
SQ SEQUENCE 144 AA; 15363 MW; C06025DA1B257128 CRC64;

Query Match          0.8%; Score 8; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 52 VKLITPDG 59

RESULT 36
Q97MN7
ID Q97MN7 PRELIMINARY; PRT; 146 AA.
AC Q97MN7:
DT 01-OCT-2001 (TREMBLrel. 18, Created)

```

```

DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
PTS system, mannitol-specific IIA domain (Ntr-type) (gene
DE MTF).
DE CAC0156.
GN Clostridium acetobutylicum.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Koonin E.V., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AF007528; AAK78140.1; -.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR ProDom: PD001689; PTS_EIIA_2; 1.
DR PROSITE: PS00372; PTS_EIIA_2; 1.
KW Complete proteome.
SQ SEQUENCE 146 AA; 16318 MW; A9A1C73CCD74DF97 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
Db 58 NGVAIPHG 55
|||||

RESULT 37
Q8RCS1 PRELIMINARY; PRT; 146 AA.
ID Q8RCS1
AC Q8RCS1
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Phosphotransferase system mannitol/fructose-specific IIA domain
DE (Ntr-type).
GN PTSN2 OR TPE0341.
OS Thermoanaerobacter tengcongensis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AF013006; AAM23633.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 146 AA; 16368 MW; ADB28A45BEB8835 CRC64;

```

```

Query Match 0.8%; Score 8; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
Db 58 NGVAIPHG 55
|||||

```

```

RESULT 38
Q8YEUI PRELIMINARY; PRT; 154 AA.
ID Q8YEUI
AC Q8YEUI
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Nitrogen regulatory IIA protein (EC 2.7.1.69).
GN BME11786.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009611; AAL52967.1; -.
DR InterPro: IPR004715; PTS_EIIA_2.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR ProDom: PD001689; PTS_EIIA_2; 1.
DR TIGRFAMS: TIGR00848; fruA; 1.
DR PROSITE: PS00372; PTS_EIIA_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 154 AA; 16497 MW; D53523086D684FB4 CRC64;

```

```

Query Match 0.8%; Score 8; DB 16; Length 154;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 303 NGVAIPHG 310
Db 60 NGVAIPHG 67
|||||

```

```

RESULT 39
Q9GFJO PRELIMINARY; PRT; 182 AA.
ID Q9GFJO
AC Q9GFJO
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Ribosomal protein L2 (fragment).
GN RPL2.
OS Ginkgo biloba (Ginkgo).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Ginkgoophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
OX NCBI_TaxID=3311;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms."
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL: AF123793; AAC26141.1; -.
DR HSP; P04257; IRL2.
DR InterPro: IPR002171; Ribosomal_L2.
DR Pfam: PF00181; Ribosomal_L2; 1.
DR TIGRFAMS: TIGR01171; rplB_bact; 1.
DR PROSITE: PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1 1

```

```

RT basal angiosperms."
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL: AF123791; AAC26139.1; -.
DR HSP; P04257; 1RL2.
DR InterPro: IPR002171; Ribosomal_L2.
DR Pfam: PF00181; Ribosomal_L2; 1.
DR TIGRFAMs: TIGR01171; rplB_bact; 1.
DR DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1 1
FT NON_TER 190 190
SO SEQUENCE 190 AA; 20210 MW; D51F3555C9439798 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 190;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 124 LRLPSGEV 131
|||||||

RESULT 42
Q9GF16 PRELIMINARY; PRT: 197 AA.
AC Q9GF16;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Saururus cernuus (Lizardtail).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Piperales, Saururaceae; Saururus.
OX NCBI_TaxID=13260;
([1])
RN RNN SEQUENCE FROM N.A.
RP pubMed=11080123;
RX Graham S.W., Olmstead R.G.;
RA "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms."
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL: AF123797; AAC26145.1; -.
DR HSP; P04257; 1RL2.
DR InterPro: IPR002171; Ribosomal_L2.
DR Pfam: PF00181; Ribosomal_L2; 1.
DR TIGRFAMs: TIGR01171; rplB_bact; 1.
DR DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1 1
FT NON_TER 197 197
SO SEQUENCE 197 AA; 20965 MW; 77FE203A3504949B CRC64;

Query Match 0.8%; Score 8; DB 8; Length 197;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 125 LRLPSGEV 132
|||||||

RESULT 43
Q9GFJ5 PRELIMINARY; PRT: 199 AA.
AC Q9GFJ5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Calycanthus floridus (Sweet shrub).

```

```

OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Nerales; Calycanthaceae; Calycanthus.
CX NCBI_TaxID=3429;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123788; AAG26136.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 199
SQ SEQUENCE 199 AA; 21316 MW; 2D2673BC6CF20DCD CRC64;

Query Match 0.8%; Score 8; DB 8; Length 199;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 124 LRLPSGEV 131

RESULT 44
Q9GFJ6 PRELIMINARY; PRT; 200 AA.
AC Q9GFJ6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Cabomba caroliniana.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Cabombaceae; Cabomba.
CX NCBI_TaxID=4426;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123787; AAG26135.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 200
SQ SEQUENCE 200 AA; 21334 MW; 41F8BE657EF9D54 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 200;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 126 LRLPSGEV 133

RESULT 45

```

```

Q9GFJ8 PRELIMINARY; PRT; 201 AA.
AC Q9GFJ8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Acorus calamus (Sweet flag).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Acoraceae; Acorus.
CX NCBI_TaxID=4465;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123785; AAG26133.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 201
SQ SEQUENCE 201 AA; 21349 MW; 15441A79378F5459 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 201;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 126 LRLPSGEV 133

Search completed: May 13, 2003, 14:01:05
Job time : 55 secs

```

GenCore version 5.1.4_p5_4578
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:59:18 ; Search time 19 Seconds
(without alignments)
1608.969 Million cell updates/sec

Title: US-09-471-255-2
Perfect score: 1039
Sequence: 1 MKFSKXYIAGSAVIVSLSL.....IELRLPSGEVKKNLSDFIA 1039

Scoring table: Oligo
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 8
Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : Issued Patents.AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	447	43.0	447	4	US-08-961-083-182
2	32	3.1	796	4	US-08-961-083-56
3	28	2.7	763	4	US-08-961-083-66
4	8	0.8	144	2	US-09-049-577-2
5	8	0.8	144	4	US-09-390-598-2

ALIGNMENTS

RESULT 1
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-182

Query Match 43.0%; Score 447; DB 4; Length 447;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 LNHRSSENKNNRVSYVDGSSQSKSENLTDPDQVSQKEGIAEQIVIKITDQGYVTSHG 84
DB 1 LNHRSSENKNNRVSYVDGSSQSKSENLTDPDQVSQKEGIAEQIVIKITDQGYVTSHG 60
QY 85 DHYHYNGKVPYDALFSEELLKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYLKDAHA 144
DB 61 DHYHYNGKVPYDALFSEELLKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYLKDAHA 120
QY 145 DNVRTKDEINRQKEHVKNDEKYNVAVARSGRYTTNDGYVFNPAIITDTCNAYIVP 204
DB 121 DNVRTKDEINRQKEHVKNDEKYNVAVARSGRYTTNDGYVFNPAIITDTCNAYIVP 180
QY 205 HGGHYHVIKPSDLASALAHAHLAGKNMOPSOISYSSSTASDNTQSVAKGSTSKPANK 264
DB 181 HGGHYHVIKPSDLASALAHAHLAGKNMOPSOISYSSSTASDNTQSVAKGSTSKPANK 240
QY 265 SENLQSLKELYDPSAQRYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSKLSAL 324
DB 241 SENLQSLKELYDPSAQRYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSKLSAL 300
QY 325 EEKTAHVPISTGSTVSTNAKPNENVYSSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDI 384
DB 301 EEKTAHVPISTGSTVSTNAKPNENVYSSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDI 360
QY 385 VEETATAYIVRHGDHFYIPKSNQIGOPTLPNNSLATPSPSLNPCTSHKHEEDGYGF 444
DB 361 VEETATAYIVRHGDHFYIPKSNQIGOPTLPNNSLATPSPSLNPCTSHKHEEDGYGF 420
QY 445 DANRIIAEDSGFVMSHGDHNNHYFFKK 471
DB 421 DANRIIAEDSGFVMSHGDHNNHYFFKK 447

RESULT 2

US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 3.1%; Score 32; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 7.1e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVKITDQGYVTSRHDHYHYNGKVPYDA 98
|||||
DB 46 AEQIVKITDQGYVTSRHDHYHYNGKVPYDA 77
|||||

RESULT 3
US-08-961-083-66
Sequence 66, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 2.7%; Score 28; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 6.1e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 VIKITDQGYVTSRHDHYHYNGKVPYDA 98
|||||
DB 52 VIKITDQGYVTSRHDHYHYNGKVPYDA 79
|||||

RESULT 4
US-09-049-577-2
Sequence 2, Application US/09049577
Patent No. 5968804
GENERAL INFORMATION:
APPLICANT: Feng, Teng-yung
APPLICANT: Lin, Hao-Jan
TITLE OF INVENTION: AMPHIPATHIC PROTEIN-1
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/049,577
FILING DATE: 27-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34,053
REFERENCE/DOCKET NUMBER: 08919/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-049-577-2

Query Match 0.8%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 VKLITPDG 648
|||||
DB 52 VKLITPDG 59
|||||

RESULT 5
US-09-390-598-2
; Sequence 2, Application US/09390598
; Patent No. 6297360
; GENERAL INFORMATION:
; APPLICANT: Feng, Teng-yung
; APPLICANT: Lin, Hao-jan
; TITLE OF INVENTION: AMPHIPATHIC PROTEIN-1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/390,598
; FILING DATE: 27-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34,053
; REFERENCE/DOCKET NUMBER: 08919/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-390-598-2

Query Match 0.8%; Score 8; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 52 VKLITPDG 59

Search completed: May 13, 2003, 14:01:59
Job time : 20 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 14:00:08 ; Search time 28 Seconds
(without alignments)
3414.811 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKYYIAGSAVIVLSL.....IELRLPSGEVKKNLSDPIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 349150.seqs, 92025710 residues

Word size : 8

Total number of hits satisfying chosen parameters: 77

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	1039	9	US-09-884-465A-6
2	807	77.7	1152	9	US-09-884-465A-379
3	807	77.7	1238	9	US-09-884-465A-381
4	807	77.7	1365	9	US-09-884-465A-382
5	807	77.7	1378	9	US-09-884-465A-378
6	578	55.6	633	9	US-09-884-465A-350
7	575	55.3	633	9	US-09-884-465A-351
8	575	55.3	633	9	US-09-884-465A-361
9	573	55.1	627	9	US-09-884-465A-366
10	573	55.1	633	9	US-09-884-465A-362
11	573	55.1	894	9	US-09-884-465A-337
12	573	55.1	895	9	US-09-884-465A-347
13	573	55.1	900	9	US-09-884-465A-335
14	573	55.1	900	9	US-09-884-465A-339
15	573	55.1	900	9	US-09-884-465A-341
16	573	55.1	901	9	US-09-884-465A-343
17	573	55.1	901	9	US-09-884-465A-345
18	568	54.7	569	9	US-09-884-465A-235
19	568	54.7	569	9	US-09-884-465A-255

ALIGNMENTS

RESULT 1

US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie

20	568	54.7	913	9	US-09-884-465A-384	Sequence 384, App
21	568	54.7	999	9	US-09-884-465A-376	Sequence 376, App
22	568	54.7	999	9	US-09-884-465A-377	Sequence 377, App
23	568	54.7	1126	9	US-09-884-465A-383	Sequence 383, App
24	568	54.7	1139	9	US-09-884-465A-380	Sequence 380, App
25	531	51.1	633	9	US-09-884-465A-354	Sequence 354, App
26	489	47.1	633	9	US-09-884-465A-352	Sequence 352, App
27	489	47.1	633	9	US-09-884-465A-357	Sequence 357, App
28	489	47.1	633	9	US-09-884-465A-358	Sequence 358, App
29	489	47.1	633	9	US-09-884-465A-359	Sequence 359, App
30	489	47.1	633	9	US-09-884-465A-360	Sequence 360, App
31	489	47.1	900	9	US-09-884-465A-333	Sequence 333, App
32	489	47.1	900	9	US-09-884-465A-334	Sequence 334, App
33	489	47.1	906	9	US-09-884-465A-332	Sequence 332, App
34	489	47.1	906	9	US-09-884-465A-369	Sequence 369, App
35	489	47.1	906	9	US-09-884-465A-370	Sequence 370, App
36	489	47.1	906	9	US-09-884-465A-373	Sequence 373, App
37	486	46.8	569	9	US-09-884-465A-356	Sequence 356, App
38	486	46.8	569	9	US-09-884-465A-353	Sequence 353, App
39	486	46.8	569	9	US-09-884-465A-355	Sequence 355, App
40	486	46.8	569	9	US-09-884-465A-363	Sequence 363, App
41	486	46.8	569	9	US-09-884-465A-364	Sequence 364, App
42	486	46.8	569	9	US-09-884-465A-365	Sequence 365, App
43	486	46.8	569	9	US-09-884-465A-371	Sequence 371, App
44	486	46.8	569	9	US-09-884-465A-372	Sequence 372, App
45	484	46.6	569	9	US-09-884-465A-256	Sequence 256, App
46	484	46.6	621	9	US-09-884-465A-368	Sequence 368, App
47	484	46.6	627	9	US-09-884-465A-367	Sequence 367, App
48	484	46.6	633	9	US-09-884-465A-349	Sequence 349, App
49	484	46.6	633	9	US-09-884-465A-338	Sequence 338, App
50	484	46.6	889	9	US-09-884-465A-348	Sequence 348, App
51	484	46.6	894	9	US-09-884-465A-336	Sequence 336, App
52	484	46.6	894	9	US-09-884-465A-340	Sequence 340, App
53	484	46.6	894	9	US-09-884-465A-342	Sequence 342, App
54	484	46.6	895	9	US-09-884-465A-344	Sequence 344, App
55	484	46.6	895	9	US-09-884-465A-346	Sequence 346, App
56	475	45.7	840	9	US-09-884-465A-10	Sequence 10, Appl
57	447	43.0	447	10	US-09-765-272-182	Sequence 182, App
58	94	9.0	94	9	US-09-884-465A-11	Sequence 11, Appl
59	78	7.5	78	9	US-09-884-465A-19	Sequence 19, Appl
60	57	5.5	57	9	US-09-884-465A-17	Sequence 17, Appl
61	51	4.9	51	9	US-09-884-465A-20	Sequence 20, Appl
62	42	4.0	42	9	US-09-884-465A-12	Sequence 12, Appl
63	40	3.8	40	9	US-09-884-465A-18	Sequence 18, Appl
64	39	3.8	39	9	US-09-884-465A-13	Sequence 13, Appl
65	36	3.5	36	9	US-09-884-465A-14	Sequence 14, Appl
66	35	3.4	35	9	US-09-884-465A-16	Sequence 16, Appl
67	32	3.1	796	10	US-09-765-272-56	Sequence 56, Appl
68	32	3.1	838	9	US-09-884-465A-8	Sequence 8, Appl
69	32	3.1	840	9	US-09-884-465A-7	Sequence 7, Appl
70	28	2.7	763	10	US-09-765-272-66	Sequence 66, Appl
71	24	2.3	24	9	US-09-884-465A-15	Sequence 15, Appl
72	16	1.5	793	9	US-09-252-088-15	Sequence 15, Appl
73	8	0.8	543	9	US-09-746-783-78	Sequence 78, Appl
74	8	0.8	715	9	US-09-252-088-16	Sequence 16, Appl
75	8	0.8	1134	9	US-09-836-392-16	Sequence 16, Appl
76	8	0.8	1237	10	US-09-862-027-78	Sequence 78, Appl
77	8	0.8	1308	10	US-09-862-027-79	Sequence 79, Appl

APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 100.0%; Score 1039; DB 9; Length 1039;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1039; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFSKYIAGSAVIVSLCAYALNHRSQENKNNRVSYVDGSSOSKSENLTDPQVS 60
DB 1 MKFSKYIAGSAVIVSLCAYALNHRSQENKNNRVSYVDGSSOSKSENLTDPQVS 60

QY 61 QKGGIOAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYOLKDAIVN 120
DB 61 QKGGIOAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYOLKDAIVN 120

QY 121 EVKGGYIIVKGVYVYLKDAHADNVRKDEINROKQDHVNDNEKVNNAVARSQRY 180
DB 121 EVKGGYIIVKGVYVYLKDAHADNVRKDEINROKQDHVNDNEKVNNAVARSQRY 180

QY 181 TTNDGVYVNPADIIEDTGNAVIVPHGCHHYHYIPKSDLSASELAHAHLAGNMPQSLS 240
DB 181 TTNDGVYVNPADIIEDTGNAVIVPHGCHHYHYIPKSDLSASELAHAHLAGNMPQSLS 240

QY 241 YSTASDNNTQSVAKGSTSKPANKSENLSLLKELYDPSAQRYSSESDGLVDPKIIIS 300
DB 241 YSTASDNNTQSVAKGSTSKPANKSENLSLLKELYDPSAQRYSSESDGLVDPKIIIS 300

QY 301 TPNGVAIPHGDHYHFIPIYKLSALEEKIARMVPISTGTVSTNAKPNEVWSLGSLSN 360
DB 301 TPNGVAIPHGDHYHFIPIYKLSALEEKIARMVPISTGTVSTNAKPNEVWSLGSLSN 360

QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQPTLPNNSLA 420
DB 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQPTLPNNSLA 420

QY 421 TPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHYFFPKKOLTEQIKA 480
DB 421 TPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHYFFPKKOLTEQIKA 480

QY 481 AQHLEEVKTSHNGLDLSLSSHEODYPCNAKEMKDLKKEEIKIAGIMKOYGVKRESIVVN 540
DB 481 AQHLEEVKTSHNGLDLSLSSHEODYPCNAKEMKDLKKEEIKIAGIMKOYGVKRESIVVN 540

QY 541 KEKNALIIYPHGDHHDADIDDEHPKVPVIGHSNHYELFKPEEGVAKKEGKRVITGEELTNV 600
DB 541 KEKNALIIYPHGDHHDADIDDEHPKVPVIGHSNHYELFKPEEGVAKKEGKRVITGEELTNV 600

QY 601 VNLKNTFNQNTLANGOKRYVSFPFPPLEKLGINMLVLIITPDGKVLKVSQKVF 660
DB 601 VNLKNTFNQNTLANGOKRYVSFPFPPLEKLGINMLVLIITPDGKVLKVSQKVF 660

QY 661 EGVGNIAINFELDQYLPQGTQKTYTIAASKQYPEVSYDGTFTVPTVTSLAYKVASQITIFPFFIA 720
DB 661 EGVGNIAINFELDQYLPQGTQKTYTIAASKQYPEVSYDGTFTVPTVTSLAYKVASQITIFPFFIA 720

QY 721 GDTVLRVNPQFVAPKGDALVRVDFEFGNAYLENNYKVEIKLPIPKLNOGTTTRAGNK 780
DB 721 GDTVLRVNPQFVAPKGDALVRVDFEFGNAYLENNYKVEIKLPIPKLNOGTTTRAGNK 780

QY 781 IPTVFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEPKTS 840
DB 781 IPTVFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEPKTS 840

QY 841 EKVEKEKLSSETNSTLEEVPYTPVQVEKVAKFAESYGMKLENVLFNMDGTIELYLP 900
DB 841 EKVEKEKLSSETNSTLEEVPYTPVQVEKVAKFAESYGMKLENVLFNMDGTIELYLP 900

QY 901 SCEVIKKKNMADFTGEAPQNGENKPSKNGKSTGTVENQPTENKPADSLPEAPNEKPKP 960
DB 901 SCEVIKKKNMADFTGEAPQNGENKPSKNGKSTGTVENQPTENKPADSLPEAPNEKPKP 960

QY 961 ENSTNGMLNPEGNVSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGDSVIFNMDGTI 1020
DB 961 ENSTNGMLNPEGNVSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGDSVIFNMDGTI 1020

QY 1021 ELRLPSGEVIVKKNLSDFIA 1039
DB 1021 ELRLPSGEVIVKKNLSDFIA 1039

RESULT 2
US-09-884-465A-379
; Sequence 379, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 379
; LENGTH: 1152
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (344)..(344)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-379

Query Match 77.7%; Score 807; DB 9; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLLKELYDPSAQRYSSESDGLVF 292
DB 346 NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLLKELYDPSAQRYSSESDGLVF 405

QY 293 DPAKIIISRTPNGVAIPHGDHYHFIPIYKLSALEEKIARMVPISTGTVSTNAKPNEVWS 352
DB 406 DPAKIIISRTPNGVAIPHGDHYHFIPIYKLSALEEKIARMVPISTGTVSTNAKPNEVWS 465

QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQ 412
DB 466 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQ 525

QY 413 TLPNNSLATPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHYFFK 472
DB 525 TLPNNSLATPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHYFFK 472

Db 526 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 585
QY 473 LTEEQIKAAQKHLLEEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIAGIMKQYGV 532
Db 586 LTEEQIKAAQKHLLEEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIAGIMKQYGV 645
QY 533 KRESIVVNEKNAIIPHGDDHHDIDDEKPVGIGHSHSNVELPKPEGVAKKGNKVY 592
Db 646 KRESIVVNEKNAIIPHGDDHHDIDDEKPVGIGHSHSNVELPKPEGVAKKGNKVY 705
QY 593 TGEELNVVNLKNSFNNQNTLANGQKRVSFSPPELEKLGINMLVCLTTPDGKYLE 652
Db 706 TGEELNVVNLKNSFNNQNTLANGQKRVSFSPPELEKLGINMLVCLTTPDGKYLE 765
QY 653 KVSCKVFGGVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
Db 766 KVSCKVFGGVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 825
QY 713 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFPHGNAYLENNYKVGKIKLPIPKLNOG 772
Db 826 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFPHGNAYLENNYKVGKIKLPIPKLNOG 885
QY 773 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 832
Db 886 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 945
QY 833 KVEEPTSEKVEKEKLSGTGNSSTNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMD 892
Db 946 KVEEPTSEKVEKEKLSGTGNSSTNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMD 1005
QY 893 GTIELYLPSEGEVIKKNMADFTGEAQQNGENKPSNGKYSTGTGTVENQPTENKPADSLPEA 952
Db 1006 GTIELYLPSEGEVIKKNMADFTGEAQQNGENKPSNGKYSTGTGTVENQPTENKPADSLPEA 1065
QY 953 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
Db 1066 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1125
QY 1013 IFNMDGTIELRLPSEGEVIKKNLSDFIA 1039
Db 1126 IFNMDGTIELRLPSEGEVIKKNLSDFIA 1152

RESULT 3

US-09-884-465A-381
; Sequence 381, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212, 683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (430)..(430)

RESULT 4

US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.

; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-381

Query Match 77.7%; Score 807; DB 9; Length 1238;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMOPSQLSYSTASDNNTQSVAKGSTKSPANKSENLOSLLKELYDSPAQRYSDEGLVF 292

Db 432 NMOPSQLSYSTASDNNTQSVAKGSTKSPANKSENLOSLLKELYDSPAQRYSDEGLVF 491

QY 293 DPAKIIISITPNGVAIPHGDDHVFIPYSKLSALEEKIAMVPIISGTSTVSTNAKPNVVS 352

Db 492 DPAKIIISITPNGVAIPHGDDHVFIPYSKLSALEEKIAMVPIISGTSTVSTNAKPNVVS 551

QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVETATATAYIVRHGDHFIHYTPKSNQIQP 412

Db 552 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVETATATAYIVRHGDHFIHYTPKSNQIQP 611

QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 472

Db 612 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 671

QY 473 LTEEQIKAAQKHLLEEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIAGIMKQYGV 532

Db 672 LTEEQIKAAQKHLLEEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLDDKIEKIAGIMKQYGV 731

QY 533 KRESIVVNEKNAIIPHGDDHHDIDDEKPVGIGHSHSNVELPKPEGVAKKGNKVY 592

Db 732 KRESIVVNEKNAIIPHGDDHHDIDDEKPVGIGHSHSNVELPKPEGVAKKGNKVY 791

QY 593 TGEELNVVNLKNSFNNQNTLANGQKRVSFSPPELEKLGINMLVCLTTPDGKYLE 652

Db 792 TGEELNVVNLKNSFNNQNTLANGQKRVSFSPPELEKLGINMLVCLTTPDGKYLE 851

QY 653 KVSCKVFGGVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 712

Db 852 KVSCKVFGGVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 911

QY 713 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFPHGNAYLENNYKVGKIKLPIPKLNOG 772

Db 912 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFPHGNAYLENNYKVGKIKLPIPKLNOG 971

QY 773 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 832

Db 972 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 1031

QY 833 KVEEPTSEKVEKEKLSGTGNSSTNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMD 892

Db 1032 KVEEPTSEKVEKEKLSGTGNSSTNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMD 1091

QY 893 GTIELYLPSEGEVIKKNMADFTGEAQQNGENKPSNGKYSTGTGTVENQPTENKPADSLPEA 952

Db 1092 GTIELYLPSEGEVIKKNMADFTGEAQQNGENKPSNGKYSTGTGTVENQPTENKPADSLPEA 1151

QY 953 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012

Db 1152 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1211

QY 1013 IFNMDGTIELRLPSEGEVIKKNLSDFIA 1039

Db 1212 IFNMDGTIELRLPSEGEVIKKNLSDFIA 1238

APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 382
LENGTH: 1365
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (557)..(557)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (558)..(558)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-382

Query Match 77.7%; Score 807; DB 9; Length 1365;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESGLVF 292
DB 559 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESGLVF 618
QY 293 DPAKIIERTPNGVAIPGHDHYHFIPIYSKLSALEEKTARMPVIGTGSTVSTNAKNEVVS 352
DB 619 DPAKIIERTPNGVAIPGHDHYHFIPIYSKLSALEEKTARMPVIGTGSTVSTNAKNEVVS 678
QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIPIKSNQIGQP 412
DB 679 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIPIKSNQIGQP 738
QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 472
DB 739 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 798
QY 473 LTERQIRAAQKHLDEVKTSINGLSDLSSSHEDQYPGNAKEMKOLDKIEEKIAGIMKQYGV 532
DB 799 LTERQIRAAQKHLDEVKTSINGLSDLSSSHEDQYPGNAKEMKOLDKIEEKIAGIMKQYGV 858
QY 533 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNVELFPEGVAKKEGNKYV 592
DB 859 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNVELFPEGVAKKEGNKYV 918
QY 593 TGEELTNVLLKNSTNNQNTLANGQKRVSFSPFPELEKKGILGNLKLITPDGKYLE 652
DB 919 TGEELTNVLLKNSTNNQNTLANGQKRVSFSPFPELEKKGILGNLKLITPDGKYLE 978
QY 653 KVSQKVGEGVGNITANFELDQPLPGQTFYTTASKDYFEVSDGTFYPTSLAYKMASQ 712
DB 979 KVSQKVGEGVGNITANFELDQPLPGQTFYTTASKDYFEVSDGTFYPTSLAYKMASQ 1038
QY 713 TIFYPFHAGDTYLRVNPQFAPVPGTDAIVRVDFEFGNAYLENNYKVGIEKIPKPLNQG 772
DB 1039 TIFYPFHAGDTYLRVNPQFAPVPGTDAIVRVDFEFGNAYLENNYKVGIEKIPKPLNQG 1098
QY 773 TTRTAGNKIPVTFMANAYLONQSTYIVVEPILEKNOTKPSILPOLFKKNAQENSKLDE 832
DB 1099 TTRTAGNKIPVTFMANAYLONQSTYIVVEPILEKNOTKPSILPOLFKKNAQENSKLDE 1158

QY 833 KVEEPKTSKVEKEKLETSNSTNSTLEEVPTVDVQEKVAKFAESYGMKLENVLFNMD 892
DB 1159 KVEEPKTSKVEKEKLETSNSTNSTLEEVPTVDVQEKVAKFAESYGMKLENVLFNMD 1218
QY 893 GTIELYLPSPGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGVENQPTENKPADSLPEA 952
DB 1219 GTIELYLPSPGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGVENQPTENKPADSLPEA 1278
QY 953 PNEKPVKPNSTNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
DB 1279 PNEKPVKPNSTNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1338
QY 1013 IFNMDGTIELRLPSGEVIKKNLSDFIA 1039
DB 1339 IFNMDGTIELRLPSGEVIKKNLSDFIA 1365

RESULT 5

US-09-884-465A-378
Sequence 378, Application US/09884465A
Publication No US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 378
LENGTH: 1378
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (570)..(570)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (571)..(571)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-378

Query Match 77.7%; Score 807; DB 9; Length 1378;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESGLVF 292
DB 572 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESGLVF 631
QY 293 DPAKIIERTPNGVAIPGHDHYHFIPIYSKLSALEEKTARMPVIGTGSTVSTNAKNEVVS 352
DB 632 DPAKIIERTPNGVAIPGHDHYHFIPIYSKLSALEEKTARMPVIGTGSTVSTNAKNEVVS 691
QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIPIKSNQIGQP 412
DB 692 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIPIKSNQIGQP 751
QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 472
DB 752 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 811

QY 473 LTEQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKDLKKIEKIMQYGV 532
DB 812 LTEQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKDLKKIEKIMQYGV 871
QY 533 KRESIVVAKENXNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVY 592
DB 872 KRESIVVAKENXNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVY 931
QY 593 TGEELTNVNLKNSFTNNQNTLANGOKRVSFSPFPPELEKLGINMLVKLITPDGKYLE 652
DB 932 TGEELTNVNLKNSFTNNQNTLANGOKRVSFSPFPPELEKLGINMLVKLITPDGKYLE 991
QY 653 KVSQVFGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
DB 992 KVSQVFGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 1051
QY 713 TIFYPFHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGHNAYLNNYKVGGEIKLPKLNQ 772
DB 1052 TIFYPFHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGHNAYLNNYKVGGEIKLPKLNQ 1111
QY 773 TTRTAGNKIPVTFMANAYLDNOSTYIVVEPILKENDQTKPSILPQFKRNKAQENSKLDE 832
DB 1112 TTRTAGNKIPVTFMANAYLDNOSTYIVVEPILKENDQTKPSILPQFKRNKAQENSKLDE 1171
QY 833 KVEEPTSEKVEKEKLSGTGNSNSTLEEYPTVDPVQEKVAKFAESYGMKLENVFNMD 892
DB 1172 KVEEPTSEKVEKEKLSGTGNSNSTLEEYPTVDPVQEKVAKFAESYGMKLENVFNMD 1231
QY 893 GTIELYPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEA 952
DB 1232 GTIELYPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEA 1291
QY 953 PNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSV 1012
DB 1292 PNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSV 1351
QY 1013 IFNMDGTIELRPSGVEIKKNLSDFIA 1039
DB 1352 IFNMDGTIELRPSGVEIKKNLSDFIA 1378

RESULT 6

US-09-884-465A-350
; Sequence 350, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 350
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-350

Query Match 55.6%; Score 578; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GDHNHYFFKDLTEEQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKDLKKIEE 521
DB 56 GDHNHYFFKDLTEEQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKDLKKIEE 115
QY 522 KIAGIMQYGVKRESIVVAKENXNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEE 561
DB 116 KIAGIMQYGVKRESIVVAKENXNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEE 175
QY 582 GVAKKEGNKYVTGEEELTNVNLKNSFTNNQNTLANGOKRVSFSPFPPELEKLGINMLV 641
DB 176 GVAKKEGNKYVTGEEELTNVNLKNSFTNNQNTLANGOKRVSFSPFPPELEKLGINMLV 235
QY 642 KLITPDGKYLEKVSQVFGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTV 701
DB 236 KLITPDGKYLEKVSQVFGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTV 295
QY 702 PTLAYKMASQTIIFYPHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGHNAYLNNYKVG 761
DB 296 PTLAYKMASQTIIFYPHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGHNAYLNNYKVG 355
QY 762 IKPIPLNGGTRTAGNKIPVTFMANAYLDNOSTYIVVEPILKENDQTKPSILPQFKR 821
DB 356 IKPIPLNGGTRTAGNKIPVTFMANAYLDNOSTYIVVEPILKENDQTKPSILPQFKR 415
QY 822 NKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEYPTVDPVQEKVAKFAESY 881
DB 416 NKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEYPTVDPVQEKVAKFAESY 475
QY 882 MKLENVFNMDGTIELYPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTVENQPT 941
DB 476 MKLENVFNMDGTIELYPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTVENQPT 535
QY 942 ENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEK 1001
DB 536 ENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEK 595
QY 1002 TASYGLGDSVIFNMDGTIELRPSGVEIKKNLSDFIA 1039
DB 596 TASYGLGDSVIFNMDGTIELRPSGVEIKKNLSDFIA 633

RESULT 7

US-09-884-465A-351
; Sequence 351, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 351
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-351

Query Match 55.3%; Score 575; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEEQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKDLKKIEKIA 524

Db 59 NHYFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIA 118
QY 525 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDHKKPVGIGHSHSNYELFKPEEGVA 584
Db 119 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDHKKPVGIGHSHSNYELFKPEEGVA 178
QY 585 KKEGNKYVTGEELTNVNLKSNFTNNQNTLANGORVRSFSPPELEKLGINMLVKLI 644
Db 179 KKEGNKYVTGEELTNVNLKSNFTNNQNTLANGORVRSFSPPELEKLGINMLVKLI 238
QY 645 TPDCKVLEKSVKGVGEGVGNANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 704
Db 239 TPDCKVLEKSVKGVGEGVGNANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 298
QY 705 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVFDEFGHGNAYLENNKYVGEIKL 764
Db 299 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVFDEFGHGNAYLENNKYVGEIKL 358
QY 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIYVEVPILEKENQTDKPSILPQFKRKA 824
Db 359 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIYVEVPILEKENQTDKPSILPQFKRKA 418
QY 825 QENSKLDERVEEPTSEKVEKEKLSSTNSTLEEVPTVDPVQEKVAKFAESYGMKL 884
Db 419 QENSKLDERVEEPTSEKVEKEKLSSTNSTLEEVPTVDPVQEKVAKFAESYGMKL 478
QY 885 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSTGTVENQPTENK 944
Db 479 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSTGTVENQPTENK 538
QY 945 PADSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 539 PADSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 598
QY 1005 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 599 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 633

RESULT 8

US-09-884-465A-361
; Sequence 361, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 361
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-361

Query Match 55.3%; Score 575; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIA 524
Db 55 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIAGI 114

Db 59 NHYFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIA 118
QY 525 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDHKKPVGIGHSHSNYELFKPEEGVA 584
Db 119 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDHKKPVGIGHSHSNYELFKPEEGVA 178
QY 585 KKEGNKYVTGEELTNVNLKSNFTNNQNTLANGORVRSFSPPELEKLGINMLVKLI 644
Db 179 KKEGNKYVTGEELTNVNLKSNFTNNQNTLANGORVRSFSPPELEKLGINMLVKLI 238
QY 645 TPDCKVLEKSVKGVGEGVGNANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 704
Db 239 TPDCKVLEKSVKGVGEGVGNANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 298
QY 705 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVFDEFGHGNAYLENNKYVGEIKL 764
Db 299 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVFDEFGHGNAYLENNKYVGEIKL 358
QY 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIYVEVPILEKENQTDKPSILPQFKRKA 824
Db 359 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIYVEVPILEKENQTDKPSILPQFKRKA 418
QY 825 QENSKLDERVEEPTSEKVEKEKLSSTNSTLEEVPTVDPVQEKVAKFAESYGMKL 884
Db 419 QENSKLDERVEEPTSEKVEKEKLSSTNSTLEEVPTVDPVQEKVAKFAESYGMKL 478
QY 885 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSTGTVENQPTENK 944
Db 479 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSTGTVENQPTENK 538
QY 945 PADSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 539 PADSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 598
QY 1005 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 599 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 633

RESULT 9

US-09-884-465A-366
; Sequence 366, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 366
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-366

Query Match 55.1%; Score 573; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIAGI 526
Db 55 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKIEEKIAGI 114

QY 527 MKOYGVKRESIVVNKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPPEGVAKK 586
DB 115 MKOYGVKRESIVVNKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPPEGVAKK 174
QY 587 EGNKYVTGEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 646
DB 175 EGNKYVTGEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 234
QY 647 DGKVLKSVGKVFEGVGNIANFELDQPYLPQGTFFKVTIASKDYPEVSYDGTFTVPTSLA 706
DB 235 DGKVLKSVGKVFEGVGNIANFELDQPYLPQGTFFKVTIASKDYPEVSYDGTFTVPTSLA 294
QY 707 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
DB 295 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 354
QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQKRNKQAE 826
DB 355 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQKRNKQAE 414
QY 827 NSKLDERVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 415 NSKLDERVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 474
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 946
DB 475 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 534
QY 947 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 535 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYG 594
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 595 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 627

RESULT 10

US-09-884-465A-362
; Sequence 362, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 362
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-362

Query Match 55.1%; Score 573; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAGI 526
DB 61 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAGI 120

QY 527 MKOYGVKRESIVVNKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPPEGVAKK 586
DB 121 MKOYGVKRESIVVNKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPPEGVAKK 180
QY 587 EGNKYVTGEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 646
DB 181 EGNKYVTGEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 240
QY 647 DGKVLKSVGKVFEGVGNIANFELDQPYLPQGTFFKVTIASKDYPEVSYDGTFTVPTSLA 706
DB 241 DGKVLKSVGKVFEGVGNIANFELDQPYLPQGTFFKVTIASKDYPEVSYDGTFTVPTSLA 300
QY 707 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
DB 301 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 360
QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQKRNKQAE 826
DB 361 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQKRNKQAE 420
QY 827 NSKLDERVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 421 NSKLDERVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 480
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 946
DB 481 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 540
QY 947 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 541 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYG 600
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 601 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 633

RESULT 11

US-09-884-465A-337
; Sequence 337, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 337
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-337

Query Match 55.1%; Score 573; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAGI 526
DB 322 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAGI 381
QY 527 MKOYGVKRESIVVNKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPPEGVAKK 586

Db 382 MKQGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 441
QY 587 EGNKYVTGEBELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGINMLVKLITP 646
Db 442 EGNKYVTGEBELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGINMLVKLITP 501
QY 647 DGKLVKSVKGVGEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGTFTVPTSLA 706
Db 502 DGKLVKSVKGVGEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGTFTVPTSLA 561
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGEIKLPI 766
Db 562 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGEIKLPI 621
QY 767 PKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 826
Db 622 PKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 681
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSLTBEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 682 NSKLDEKVEEPTSEKVEKEKLSGTNSTSLTBEVPTVDPVQEKVAKFAESYGMKLEN 741
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSKNGKYSTGTVENQPTENKPA 946
Db 742 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSKNGKYSTGTVENQPTENKPA 801
QY 947 DSLPEAPNEKPKVPENSTDNMNLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 802 DSLPEAPNEKPKVPENSTDNMNLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 861
QY 1007 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 862 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 894

RESULT 12
US-09-884-465A-347
; Sequence 347, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 347
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-347

Query Match 55.1%; Score 573; DB 9; Length 895;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEEKIAGI 526
Db 323 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEEKIAGI 382
QY 527 MKQGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 586

Db 383 MKQGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 442
QY 587 EGNKYVTGEBELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGINMLVKLITP 646
Db 443 EGNKYVTGEBELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGINMLVKLITP 502
QY 647 DGKLVKSVKGVGEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGTFTVPTSLA 706
Db 503 DGKLVKSVKGVGEGVGNIANFELDQPLPGOTFKYTIASKDYPEVSDGTFTVPTSLA 562
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGEIKLPI 766
Db 563 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGEIKLPI 622
QY 767 PKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 826
Db 623 PKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEPPILEKENQTDKPSILPQFKRKAQE 682
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSLTBEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 683 NSKLDEKVEEPTSEKVEKEKLSGTNSTSLTBEVPTVDPVQEKVAKFAESYGMKLEN 742
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSKNGKYSTGTVENQPTENKPA 946
Db 743 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSKNGKYSTGTVENQPTENKPA 802
QY 947 DSLPEAPNEKPKVPENSTDNMNLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 803 DSLPEAPNEKPKVPENSTDNMNLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 862
QY 1007 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 863 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 895

RESULT 13
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 335
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-335

Query Match 55.1%; Score 573; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEEKIAGI 526
Db 328 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDDKIEEKIAGI 387
QY 527 MKQGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 586
Db 388 MKQGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKK 447

Qy	587	EGNKVYTGEBLTNVNVLKKNSTENNQNFTLANGQKQVSFFPELEKKLGINMLVKLTTP	646
Db	448	EGNKVYTGEBLTNVNVLKKNSTENNQNFTLANGQKQVSFFPELEKKLGINMLVKLTTP	507
Qy	647	DGKVLKVSQKVFGEQVGNIANFELDQPYLPQGTQKTYTIAASDYPEVSYDGTFTVPTSLA	706
Db	508	DGKVLKVSQKVFGEQVGNIANFELDQPYLPQGTQKTYTIAASDYPEVSYDGTFTVPTSLA	567
Qy	707	YKMASQIIFYPEHAGDTYLRVNPOQFAVPKGTDALVRVDEFHGNAYLNNKVGEIKLPI	766
Db	568	YKMASQIIFYPEHAGDTYLRVNPOQFAVPKGTDALVRVDEFHGNAYLNNKVGEIKLPI	627
Qy	767	PKLNOGTRTAGNKKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPOFKRNKAQE	826
Db	628	PKLNOGTRTAGNKKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPOFKRNKAQE	687
Qy	827	NSKLDKRVBPKTSEKVEKEKISETGNSNSTLEEVPTVDVPQKVKAFKESVGMKLEN	886
Db	688	NSKLDKRVBPKTSEKVEKEKISETGNSNSTLEEVPTVDVPQKVKAFKESVGMKLEN	747
Qy	887	VLFNMDGTIELYLPSEGVTKKNMADTGEAPQNGENKPSSENGKVSTGTVENQNTPKA	946
Db	748	VLFNMDGTIELYLPSEGVTKKNMADTGEAPQNGENKPSSENGKVSTGTVENQNTPKA	807
Qy	947	DSLPEAPNEKPVKPENSTDNGMLNPGENVGSDPMLDPALEAPADVPOEKLKFTASVG	1006
Db	808	DSLPEAPNEKPVKPENSTDNGMLNPGENVGSDPMLDPALEAPADVPOEKLKFTASVG	867
Qy	1007	LGDSVTFNMDGTIELRPSGEVKKNLSDFTA	1039
Db	868	LGDSVTFNMDGTIELRPSGEVKKNLSDFTA	900

```

RESULT 14
US-09-884-465A-339
; Sequence 339, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 339
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-339

```

Qy	597	EGNKVYVYGEELTNVYNLLKNSVTENNQNFILANGCQRKVSFPPPELEKKLGINMLVKLITP	646
Db	448	EGNKVYVYGEELTNVYNLLKNSVTENNQNFILANGCQRKVSFPPPELEKKLGINMLVKLITP	507
Qy	647	DGKYLEKSVKVFEGVGNITANFELDQPYLPQGTFKYTTIASKDYEVSYDGTFTVPTSLA	706
Db	508	DGKYLEKSVKVFEGVGNITANFELDQPYLPQGTFKYTTIASKDYEVSYDGTFTVPTSLA	567
Qy	707	YKMASQITIEYFFHAGDTYLRVNPQFAPVPGTGTALYRVDFDEPHGNAYLENNYKVGEIKLPI	766
Db	568	YKMASQITIEYFFHAGDTYLRVNPQFAPVPGTGTALYRVDFDEPHGNAYLENNYKVGEIKLPI	627
Qy	767	PKLNGQTRTAGNKRIPVTFEMANAYLDNQSTIYIVEVPILLEKENQTDKPSILPQFRNKRAQE	826
Db	628	PKLNGQTRTAGNKRIPVTFEMANAYLDNQSTIYIVEVPILLEKENQTDKPSILPQFRNKRAQE	687
Qy	827	NSKLDKXVEEPPKTSSEKVEKEKLSETGNSNSTLEEVPITDVPQEKVAKFAESYGMKLEN	886
Db	688	NSKLDKXVEEPPKTSSEKVEKEKLSETGNSNSTLEEVPITDVPQEKVAKFAESYGMKLEN	747
Qy	887	VLFNMNDGTIELYLPSEGVIKKKNMADFTGEAPOGNGENKPSSENGKSVTGTVENQPTENKPA	946
Db	748	VLFNMNDGTIELYLPSEGVIKKKNMADFTGEAPOGNGENKPSSENGKSVTGTVENQPTENKPA	807
Qy	947	DSLPEASNEKVPKPENSTDNGMLNPGCNVGSDDPMLDPALEAPADVPQEKLEKFTASYG	1006
Db	808	DSLPEASNEKVPKPENSTDNGMLNPGCNVGSDDPMLDPALEAPADVPQEKLEKFTASYG	867
Qy	1007	LGLDVSIFNMNDGTIELRPSGEVVIKKNLSDFIA	1039
Db	868	LGLDVSIFNMNDGTIELRPSGEVVIKKNLSDFIA	900
RESULT 15			
US-09-884-465A-341			
; Sequence 341, Application US/09884465A			
; Publication No. US2003007293A1			
; GENERAL INFORMATION:			
; APPLICANT: Shire Biochem, Inc.			
; APPLICANT: Hamel, Josee			
; APPLICANT: Brodeur, Bernard			
; APPLICANT: Martin, Denis			
; APPLICANT: Charland, Nathalie			
; APPLICANT: Ouellet, Catherine			
; TITLE OF INVENTION: Streptococcus Antigens			
; FILE REFERENCE: 055190-0044			
; CURRENT APPLICATION NUMBER: US/09/884,465A			
; CURRENT FILING DATE: 2001-06-20			
; PRIOR APPLICATION NUMBER: 60/212,683			
; PRIOR FILING DATE: 2000-06-20			
; NUMBER OF SEQ ID NOS: 384			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 341			
; LENGTH: 900			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Unknown Organism			
US-09-884-465A-341			

Db 448 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 507
QY 647 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFFKTYTIASKDYPEVSYDGTFTVPTSLA 706
Db 509 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFFKTYTIASKDYPEVSYDGTFTVPTSLA 568
QY 707 YKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEHGNAYLENNYKVGEIKLPI 766
Db 569 YKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEHGNAYLENNYKVGEIKLPI 628
QY 767 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKNKAQE 826
Db 629 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKNKAQE 688
QY 827 NSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVTPVDPVQEKVAKFAESYGMKLEN 886
Db 689 NSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVTPVDPVQEKVAKFAESYGMKLEN 848
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSVSTGTVENQPTENKPA 946
Db 749 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSVSTGTVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPENSTDNGLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 809 DSLPEAPNEKPKPENSTDNGLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 868 LGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 900

RESULT 16
US-09-884-465A-343
; Sequence 343, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 343
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

Query Match 55.1%; Score 573; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHQDYPGNAKEMKOLDKRIEIKIAGI 526
Db 329 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHQDYPGNAKEMKOLDKRIEIKIAGI 388
QY 527 MKOYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
Db 389 MKOYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 448
QY 587 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 646
Db 448 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508

Db 449 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508
QY 647 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFFKTYTIASKDYPEVSYDGTFTVPTSLA 706
Db 509 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFFKTYTIASKDYPEVSYDGTFTVPTSLA 568
QY 707 YKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEHGNAYLENNYKVGEIKLPI 766
Db 569 YKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEHGNAYLENNYKVGEIKLPI 628
QY 767 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKNKAQE 826
Db 629 PKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKNKAQE 688
QY 827 NSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVTPVDPVQEKVAKFAESYGMKLEN 886
Db 689 NSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVTPVDPVQEKVAKFAESYGMKLEN 848
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSVSTGTVENQPTENKPA 946
Db 749 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSSENGKSVSTGTVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPENSTDNGLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 809 DSLPEAPNEKPKPENSTDNGLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 869 LGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 901

RESULT 17
US-09-884-465A-345
; Sequence 345, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 345
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

Query Match 55.1%; Score 573; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHQDYPGNAKEMKOLDKRIEIKIAGI 526
Db 329 YFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHQDYPGNAKEMKOLDKRIEIKIAGI 388
QY 527 MKOYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
Db 389 MKOYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 448
QY 587 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 646
Db 449 EGNKYVTGEELTNVNLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508

QY 647 DKVLEKVGKVGEGVGNIANFELDPQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLA 706
|||||
Db 509 DGKVLKVGKVGEGVGNIANFELDPQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLA 568
QY 707 YKMASTIIFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGKIKLPI 766
|||||
Db 569 YKMASTIIFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGKIKLPI 628
QY 767 PKLNQSTTRTAGNKIPVTFMANAYLDNSTYIIVEVPILKEKQTDKPSILPQFKRNKAQE 826
|||||
Db 629 PKLNQSTTRTAGNKIPVTFMANAYLDNSTYIIVEVPILKEKQTDKPSILPQFKRNKAQE 688
QY 827 NSKLDKVEEPTSEKVEKEKLSGTNSTNSLTLEEVPTDVPQVKAFASGYGKMLN 886
|||||
Db 689 NSKLDKVEEPTSEKVEKEKLSGTNSTNSLTLEEVPTDVPQVKAFASGYGKMLN 748
QY 887 VLFNMDGTIELYLPSEGEVKKMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPA 946
|||||
Db 749 VLFNMDGTIELYLPSEGEVKKMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPNSNDGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
|||||
Db - 809 DSLPEAPNEKPKPNSNDGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGLSDSVFNMDGTIELRLPSGEVKKNLSDFIA 1039
|||||
Db 869 LGLSDSVFNMDGTIELRLPSGEVKKNLSDFIA 901

RESULT 18

US-09-884-465A-235
; Sequence 235, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 235
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-235

Query Match 54.7%; Score 568; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIKAAQKHLEEVKTSNGLDLSSSHQDYPGNAKEMKDLDDKTEETAGIMKQYG 531
|||||
Db 2 DLTEEQIKAAQKHLEEVKTSNGLDLSSSHQDYPGNAKEMKDLDDKTEETAGIMKQYG 61
QY 532 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591
|||||
Db 62 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 121
QY 592 YTGELTNVNLKNSTFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
|||||
Db 122 YTGELTNVNLKNSTFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 181

QY 652 EKVSGKVGEGVGNIANFELDPQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMAS 711
|||||
Db 182 EKVSGKVGEGVGNIANFELDPQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMAS 241
QY 712 QTIETYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGKIKLPIPKLNQ 771
|||||
Db 242 QTIETYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGKIKLPIPKLNQ 301
QY 772 GTTATAGNKIPVTFMANAYLDNSTYIIVEVPILKEKQTDKPSILPQFKRNKAQENSKLD 831
|||||
Db 302 GTTATAGNKIPVTFMANAYLDNSTYIIVEVPILKEKQTDKPSILPQFKRNKAQENSKLD 361
QY 832 EKVEEPTSEKVEKEKLSGTNSTNSLTLEEVPTDVPQVKAFASGYGKMLNVLNFM 891
|||||
Db 362 EKVEEPTSEKVEKEKLSGTNSTNSLTLEEVPTDVPQVKAFASGYGKMLNVLNFM 421
QY 892 DGTIELYLPSEGEVKKMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 951
|||||
Db 422 DGTIELYLPSEGEVKKMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 481
QY 952 APNEKPKPNSNDGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
|||||
Db 482 APNEKPKPNSNDGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 541
QY 1012 VFNMDGTIELRLPSGEVKKNLSDFIA 1039
|||||
Db 542 VFNMDGTIELRLPSGEVKKNLSDFIA 569

RESULT 19

US-09-884-465A-255
; Sequence 255, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 255
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-255

Query Match 54.7%; Score 568; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIKAAQKHLEEVKTSNGLDLSSSHQDYPGNAKEMKDLDDKTEETAGIMKQYG 531
|||||
Db 2 DLTEEQIKAAQKHLEEVKTSNGLDLSSSHQDYPGNAKEMKDLDDKTEETAGIMKQYG 61
QY 532 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591
|||||
Db 62 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 121
QY 592 YTGELTNVNLKNSTFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
|||||
Db 122 YTGELTNVNLKNSTFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 181
QY 652 EKVSGKVGEGVGNIANFELDPQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMAS 711

Db 182 EKVSQVGEVGVNANFELDQYLPQGTFFKTYTASXDYPEVSDGTFTVPTSLAYKMAS 241
Qy 712 QTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKILPDKLNQ 771
Db 242 QTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKILPDKLNQ 301
Qy 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 831
Db 302 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 361
Qy 832 EKVEEPKTSKVEKEKLETSNGTSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 362 EKVEEPKTSKVEKEKLETSNGTSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 421
Qy 892 DGTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPADSLPE 951
Db 422 DGTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPADSLPE 481
Qy 952 APNEKPVKPNSTDNMGNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
Db 482 APNEKPVKPNSTDNMGNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 541
Qy 1012 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
Db 542 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 569
RESULT 20
US-09-884-465A-384
; Sequence 384, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 384
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (344)..(344)
; OTHER INFORMATION: xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: xaa = Proline or nothing
US-09-884-465A-384
Query Match 54.7%; Score 568; DB 9; Length 913;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 472 DLTEEQIKAAQKHLEEVKTSNGLDLSLSHSHEQDYPGNAKEMKDLDDKIEKIKAGIMKQYG 531
Db 346 DLTEEQIKAAQKHLEEVKTSNGLDLSLSHSHEQDYPGNAKEMKDLDDKIEKIKAGIMKQYG 405
Qy 532 VKRESIVVNKKNALIIYPHGDDHHPIDSEKPVGIGHSHSNVELFKPDEGVAKKGNKV 591

Db 406 VKRESIVVNKKNALIIYPHGDDHHPIDSEKPVGIGHSHSNVELFKPDEGVAKKGNKV 465
Qy 592 YTGEELTNVNLKNTFNQNFLLANGOKRVSFSEPPPELEKKGILNMLVKLITPDGKVL 651
Db 466 YTGEELTNVNLKNTFNQNFLLANGOKRVSFSEPPPELEKKGILNMLVKLITPDGKVL 525
Qy 652 EKVSQVGEVGVNANFELDQYLPQGTFFKTYTASXDYPEVSDGTFTVPTSLAYKMAS 711
Db 526 EKVSQVGEVGVNANFELDQYLPQGTFFKTYTASXDYPEVSDGTFTVPTSLAYKMAS 585
Qy 712 QTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKILPDKLNQ 771
Db 586 QTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKILPDKLNQ 645
Qy 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 831
Db 646 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 705
Qy 832 EKVEEPKTSKVEKEKLETSNGTSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 706 EKVEEPKTSKVEKEKLETSNGTSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 765
Qy 892 DGTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPADSLPE 951
Db 766 DGTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPADSLPE 825
Qy 952 APNEKPVKPNSTDNMGNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
Db 826 APNEKPVKPNSTDNMGNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 885
Qy 1012 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
Db 886 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 913
RESULT 21
US-09-884-465A-376
; Sequence 376, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 376
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: xaa = Proline or nothing
US-09-884-465A-376
Query Match 54.7%; Score 568; DB 9; Length 999;

	Best Local Similarity	100.0%;	Pred. No. 0;	
	Matches	568; Conservative	0; Mismatches	0; Indels
Qy	472	DLTEEQIAKAQKHL	EEYKTSHG	GLSDLS
		SSHEQDQPCNA	KEMDLDK	KIEEKTAGIMKQYG
		531		
Db	432	DLTEEQIAKAQKHL	EEYKTSHG	GLSDLS
		SSHEQDQPCNA	KEMDLDK	KIEEKTAGIMKQYG
		491		
Qy	532	VKRSIVYVNEKNA	IIPYHGDH	HHADPID
		IDEHKVPVGTG	HSHSNVEL	FKPPEGVAKKEGKV
		591		
Db	492	VKRSIVYVNEKNA	IIPYHGDH	HHADPID
		IDEHKVPVGTG	HSHSNVEL	FKPPEGVAKKEGKV
		551		
Qy	592	YTGELINVVNLLK	NSTFNQNT	FLANGOKRV
		SFSPPELEK	KLGINMLVK	LITPDGKVL
		651		
Db	552	YTGELINVVNLLK	NSTFNQNT	FLANGOKRV
		SFSPPELEK	KLGINMLVK	LITPDGKVL
		611		
Qy	652	EKVSCKVFGEGV	GNIANFEL	DQYLPQGT
		FKYTIASKD	YPEVSYDGT	TFVPTSLAYKNAS
		711		
Db	612	EKVSCKVFGEGV	GNIANFEL	DQYLPQGT
		FKYTIASKD	YPEVSYDGT	TFVPTSLAYKNAS
		671		
Qy	712	QTIIFYPFHAGT	YLVRNPQ	FAVPGKTD
		ALVRVDFEH	GNAYLENN	KVGEIKLPTPKLQ
		771		
Db	672	QTIIFYPFHAGT	YLVRNPQ	FAVPGKTD
		ALVRVDFEH	GNAYLENN	KVGEIKLPTPKLQ
		731		
Qy	772	GTTTATGANK	IPVTFMANAY	LDNOSTYI
		VEVPILEK	ENQTDKPS	ILPQFKRNKAQENSKLD
		831		
Db	732	GTTTATGANK	IPVTFMANAY	LDNOSTYI
		VEVPILEK	ENQTDKPS	ILPQFKRNKAQENSKLD
		791		
Qy	832	EKVEEPTSEK	VEKELSET	SGNSTSTN
		STLEEVT	PDQVEK	VAKFAESYGMKLENVLFNM
		891		
Db	792	EKVEEPTSEK	VEKELSET	SGNSTSTN
		STLEEVT	PDQVEK	VAKFAESYGMKLENVLFNM
		851		
Qy	892	DGTIELILP	SGVEIKKN	MAOFTGEA
		POGNGENK	PSNGKVS	TGTVENQPTENKPADSLPE
		951		
Db	852	DGTIELILP	SGVEIKKN	MAOFTGEA
		POGNGENK	PSNGKVS	TGTVENQPTENKPADSLPE
		911		
Qy	952	APNEKPKVP	PNSTNDGMLN	PEGNVSDP
		MLDPALEA	PADVPQ	EKLEKFTASYGLGLDS
		1011		
Db	912	APNEKPKVP	PNSTNDGMLN	PEGNVSDP
		MLDPALEA	PADVPQ	EKLEKFTASYGLGLDS
		971		
Qy	1012	VIFNM	DGTIELILP	SGVEIKKNLSDFIA
		1039		
Db	972	VIFNM	DGTIELILP	SGVEIKKNLSDFIA
		999		

QY	952	APNEKPVKPNSTDN	GMLNPEGNVSGD	PMLDP	PALEEA	PAVDPV	QVOKLEKT	TASVGL	GLDS	1011
Db	482	APNEKPVKPNSTDN	GMLNPEGNVSGD <td>PMLDP</td> <td>PALEEA <td>PAVDPV <td>QVOKLEKT <td>TASVGL <td>GLDS</td> <td>541</td> </td></td></td></td>	PMLDP	PALEEA <td>PAVDPV <td>QVOKLEKT <td>TASVGL <td>GLDS</td> <td>541</td> </td></td></td>	PAVDPV <td>QVOKLEKT <td>TASVGL <td>GLDS</td> <td>541</td> </td></td>	QVOKLEKT <td>TASVGL <td>GLDS</td> <td>541</td> </td>	TASVGL <td>GLDS</td> <td>541</td>	GLDS	541
QY	1012	VIFNMDGTTIELR	LPSPGEVIKKNL	SDFIA	1039					
Db	542	VIFNMDGTTIELR	LPSPGEVIKKNL	SDFIA	569					
<p>RESULT 23</p> <p>US-09-884-465A-383</p> <p>; Sequence 383, Application US/09884465A</p> <p>; Publication No. US20030077293A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Shire Biochem, Inc.</p> <p>; APPLICANT: Hamel, Josee</p> <p>; APPLICANT: Brodeur, Bernard</p> <p>; APPLICANT: Martin, Denis</p> <p>; APPLICANT: Charland, Nathalie</p> <p>; APPLICANT: Ouellet, Catherine</p> <p>; TITLE OF INVENTION: Streptococcus Antigens</p> <p>; FILE REFERENCE: 055190-0044</p> <p>; CURRENT APPLICATION NUMBER: US/09/884,465A</p> <p>; CURRENT FILING DATE: 2001-06-20</p> <p>; PRIOR FILING DATE: 2000-06-20</p> <p>; NUMBER OF SEQ ID NOS: 384</p> <p>; SOFTWARE: Patent in version 3.1</p> <p>; SEQ ID NO 377</p> <p>; LENGTH: 999</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Artificial Sequence</p> <p>; FEATURE:</p> <p>; OTHER INFORMATION: Unknown Organism</p> <p>; NAME/KEY: MISC.FEATURE</p> <p>; LOCATION: (1)..(1)</p> <p>; OTHER INFORMATION: xaa = Methionine or nothing</p> <p>; NAME/KEY: MISC.FEATURE</p>										

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (557)..(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-383

Query Match      54.7%: Score 568; DB 9; Length 1126;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIAAQAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEEKIAGIMKQYG 531
Db 559 DLTEEQIAAQAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEEKIAGIMKQYG 618
QY 532 VKRESIVVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
Db 619 VKRESIVVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 678
QY 592 YTGEELTNVNLLKNSTFNQNFLLANGQKRVSFSPFPELEKLGINMLVKLITPDGKVL 651
Db 679 YTGEELTNVNLLKNSTFNQNFLLANGQKRVSFSPFPELEKLGINMLVKLITPDGKVL 738
QY 652 EKVSCKVFGEGVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
Db 739 EKVSCKVFGEGVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 798
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 771
Db 799 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 858
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSIILPOKRNKAQENSKLD 831
Db 859 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSIILPOKRNKAQENSKLD 918
QY 832 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 919 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 978
QY 892 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 951
Db 979 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 1038
QY 952 APNEKPVKPNSTNGMLNPEGNGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGDS 1011
Db 1039 APNEKPVKPNSTNGMLNPEGNGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGDS 1098
QY 1012 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 1099 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1126

RESULT 24
US-09-884-465A-380
; Sequence 380, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
```

```
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 380
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-380

Query Match      54.7%: Score 568; DB 9; Length 1139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIAAQAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEEKIAGIMKQYG 531
Db 572 DLTEEQIAAQAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEEKIAGIMKQYG 631
QY 532 VKRESIVVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
Db 632 VKRESIVVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 691
QY 592 YTGEELTNVNLLKNSTFNQNFLLANGQKRVSFSPFPELEKLGINMLVKLITPDGKVL 651
Db 692 YTGEELTNVNLLKNSTFNQNFLLANGQKRVSFSPFPELEKLGINMLVKLITPDGKVL 751
QY 652 EKVSCKVFGEGVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
Db 752 EKVSCKVFGEGVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 811
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 771
Db 812 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 871
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSIILPOKRNKAQENSKLD 831
Db 872 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSIILPOKRNKAQENSKLD 931
QY 832 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 932 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 991
QY 892 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 951
Db 992 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 1051
QY 952 APNEKPVKPNSTNGMLNPEGNGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGDS 1011
Db 1052 APNEKPVKPNSTNGMLNPEGNGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGDS 1111
QY 1012 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 1112 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1139

RESULT 25
US-09-884-465A-354
; Sequence 354, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
```

; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 354
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-354

Query Match 51.1%; Score 531; DB 9; Length 633;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 631; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 408 QIGOPTLPNNSLATPSLPINPGTSHKHEEDCYGFDANRIAEDESFGVMSHGDNHY 467
Db 2 QIGOPTLPNNSLATPSLPINPGTSHKHEEDCYGFDANRIAEDESFGVMSHGDNHY 61
QY 468 FFKKDLTEEQIAAKQKLEEVKTSNGLDLSLSSHEQDPYGNKAKEMKDKKKIEKIAIGM 527
Db 62 FFKKDLTEEQIAAKQKLEEVKTSNGLDLSLSSHEQDPYGNKAKEMKDKKKIEKIAIGM 121
QY 528 KOYGKRESIVVNKEKNAILIYPHGDHHDPIDEHKPVGIGHSHSNYELPKPEGVAKKE 587
Db 122 KOYGKRESIVVNKEKNAILIYPHGDHHDPIDEHKPVGIGHSHSNYELPKPEGVAKKE 181
QY 588 GNKYTGEBELTNVYNLLKNSTFNQNETLANGQKRVSFSPPELEKLGINMLVKLITPD 647
Db 182 GNKYTGEBELTNVYNLLKNSTFNQNETLANGQKRVSFSPPELEKLGINMLVKLITPD 241
QY 648 GKVLKYSKGVFGGVGNIANFELDQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAY 707
Db 242 GKVLKYSKGVFGGVGNIANFELDQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAY 301
QY 708 KMASQTIFFYPFHAGDTYLVRVNPQFAVPRGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 767
Db 302 KMASQTIFFYPFHAGDTYLVRVNPQFAVPRGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 361
QY 768 KLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKNQTDKPSILPQFKRNKAQEN 827
Db 362 KLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKNQTDKPSILPQFKRNKAQEN 421
QY 828 SKLDEKVEEPTSEKVEKEKLSNSTNSTLEEVPTDVPQEKVAKFAESYGMKLENY 887
Db 422 SKLDEKVEEPTSEKVEKEKLSNSTNSTLEEVPTDVPQEKVAKFAESYGMKLENY 481
QY 888 LFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPAD 947
Db 482 LFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPAD 541
QY 948 SLPEAPNEKPVKPNSTNDGMLNPEGVNVDPMLDPALEAPAVDPVQEKLEKFTASYGL 1007
Db 542 SLPEAPNEKPVKPNSTNDGMLNPEGVNVDPMLDPALEAPAVDPVQEKLEKFTASYGL 601
QY 1008 GLDSVIFNMDGTIELRLPSEGVIKKNLSDFIA 1039
Db 602 GLDSVIFNMDGTIELRLPSEGVIKKNLSDFIA 633

RESULT 26
US-09-884-465A-352

; Sequence 352, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 352
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-352

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHEHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYTGEBELTNVYNLLKNSTFN 610
Db 145 GDHEHADPIDDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYTGEBELTNVYNLLKNSTFN 204
QY 611 NQNTFLANGQKRVSFSPPELEKLGINMLVKLITPDGKVLKYSKGVFGGVGNIANEE 670
Db 205 NQNTFLANGQKRVSFSPPELEKLGINMLVKLITPDGKVLKYSKGVFGGVGNIANEE 264
QY 671 LDQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRVNPQ 730
Db 265 LDQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRVNPQ 324
QY 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
Db 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNQSYIIVEPILKEKNQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
Db 385 LDNQSYIIVEPILKEKNQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMA 910
Db 445 TGNSTNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMA 504
QY 911 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLN 970
Db 505 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLN 564
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEGEVI 1030
Db 565 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEGEVI 624
QY 1031 KKNLSDFIA 1039
Db 625 KKNLSDFIA 633

RESULT 27
US-09-884-465A-357
; Sequence 357, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard.
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 357
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-357

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHHADPTDEHKPVGIGHSHSNVELFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 610
DB 145 GDHHADPTDEHKPVGIGHSHSNVELFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVGKVGEGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVGKVGEGVGNIANFE 264
QY 671 LDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQ 730
DB 265 LDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQ 324
QY 731 FAVPKGTALVRVDFEFGHGNAYLENNYKVGEIKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTALVRVDFEFGHGNAYLENNYKVGEIKLPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDQSTYIYVEVPILEKENQDTPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 385 LDQSTYIYVEVPILEKENQDTPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSLTLEEVPTDVPQEVKAFASYGKMLENVLFNMDGTIELYLPSSGEVIKKNMA 910
DB 445 TGNSTNSLTLEEVPTDVPQEVKAFASYGKMLENVLFNMDGTIELYLPSSGEVIKKNMA 504
QY 911 DFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 970
DB 505 DFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 564
QY 971 PEGNVGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 565 PEGNVGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 28
US-09-884-465A-358
; Sequence 358, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 358
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-358

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHHADPTDEHKPVGIGHSHSNVELFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 610
DB 145 GDHHADPTDEHKPVGIGHSHSNVELFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVGKVGEGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVGKVGEGVGNIANFE 264
QY 671 LDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQ 730
DB 265 LDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQ 324
QY 731 FAVPKGTALVRVDFEFGHGNAYLENNYKVGEIKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTALVRVDFEFGHGNAYLENNYKVGEIKLPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDQSTYIYVEVPILEKENQDTPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 385 LDQSTYIYVEVPILEKENQDTPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSLTLEEVPTDVPQEVKAFASYGKMLENVLFNMDGTIELYLPSSGEVIKKNMA 910
DB 445 TGNSTNSLTLEEVPTDVPQEVKAFASYGKMLENVLFNMDGTIELYLPSSGEVIKKNMA 504
QY 911 DFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 970
DB 505 DFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 564
QY 971 PEGNVGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 565 PEGNVGSDPMDLPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 29
US-09-884-465A-359
; Sequence 359, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 359
; LENGTH: 633
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-359

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHADPIDDEHKPVGIGHSHSNVLELKPPEGVAKEGKKNVYTGEEELTNVNVLLKNSTEN 610
DB 145 GDHHADPIDDEHKPVGIGHSHSNVLELKPPEGVAKEGKKNVYTGEEELTNVNVLLKNSTEN 204
QY 611 NONFTLANGQKRVSPFPPELEKKGINMLVKLITPDGKVLEKSGKVFGEVGNIANFE 670
DB 205 NONFTLANGQKRVSPFPPELEKKGINMLVKLITPDGKVLEKSGKVFGEVGNIANFE 264
QY 671 LDQPYLPGQTKFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 730
DB 265 LDQPYLPGQTKFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDFDEHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDFDEHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLE 850
DB 385 LDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLE 444
QY 851 TGNSTNSLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKNMA 910
DB 445 TGNSTNSLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKNMA 504
QY 911 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLN 970
DB 505 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLN 564
QY 971 PEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 565 PEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 30
US-09-884-465A-360
; Sequence 360, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; LENGTH: 633
; TYPE: PRI

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-360

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHADPIDDEHKPVGIGHSHSNVLELKPPEGVAKEGKKNVYTGEEELTNVNVLLKNSTEN 610
DB 145 GDHHADPIDDEHKPVGIGHSHSNVLELKPPEGVAKEGKKNVYTGEEELTNVNVLLKNSTEN 204
QY 611 NONFTLANGQKRVSPFPPELEKKGINMLVKLITPDGKVLEKSGKVFGEVGNIANFE 670
DB 205 NONFTLANGQKRVSPFPPELEKKGINMLVKLITPDGKVLEKSGKVFGEVGNIANFE 264
QY 671 LDQPYLPGQTKFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 730
DB 265 LDQPYLPGQTKFYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDFDEHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDFDEHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLE 850
DB 385 LDNQSTYIVPEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLE 444
QY 851 TGNSTNSLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKNMA 910
DB 445 TGNSTNSLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKNMA 504
QY 911 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLN 970
DB 505 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLN 564
QY 971 PEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 565 PEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 31
US-09-884-465A-333
; Sequence 333, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 333
; LENGTH: 900
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-333

Query Match 47.1%; Score 489; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 610
DB 145 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVGKVGEGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVGKVGEGVGNIANFE 264
QY 671 LDQYLPFGQTFKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 265 LDQYLPFGQTFKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNSTYIIVEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 850
DB 385 LDNSTYIIVEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 444
QY 851 TGNSTNSLIEEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 910
DB 445 TGNSTNSLIEEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 504
QY 911 DFTGEAPOGNGENKPSNGKSVGTGTVENQPTENKPADSLPEAPNEKPKPENSTNGMLN 970
DB 505 DFTGEAPOGNGENKPSNGKSVGTGTVENQPTENKPADSLPEAPNEKPKPENSTNGMLN 564
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 565 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 32
US-09-884-465A-334
; Sequence 334, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 334
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-334

Query Match 47.1%; Score 489; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 610

DB 412 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 471
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVGKVGEGVGNIANFE 670
DB 472 NQFTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVGKVGEGVGNIANFE 531
QY 671 LDQYLPFGQTFKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 532 LDQYLPFGQTFKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 591
QY 731 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 592 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 651
QY 791 LDNSTYIIVEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 850
DB 652 LDNSTYIIVEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 711
QY 851 TGNSTNSLIEEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 910
DB 712 TGNSTNSLIEEVPILEKENQDKPSILPOFKRKAQENSKLDEKVEEPKTSKVEKEKLS 771
QY 911 DFTGEAPOGNGENKPSNGKSVGTGTVENQPTENKPADSLPEAPNEKPKPENSTNGMLN 970
DB 772 DFTGEAPOGNGENKPSNGKSVGTGTVENQPTENKPADSLPEAPNEKPKPENSTNGMLN 831
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
DB 832 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 891
QY 1031 KKNLSDFIA 1039
DB 892 KKNLSDFIA 900

RESULT 33
US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332

Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 610
DB 418 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKGKGNKYTGELTNVNLKNSTFN 477
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVGKVGEGVGNIANFE 670

Db 478 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGNIANFE 537
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 730
Db 538 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 597
Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 598 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 657
Qy 791 LDQSTYIIVEVPILEKENQTKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLS 850
Db 658 LDQSTYIIVEVPILEKENQTKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLS 717
Qy 851 TGNSTNSNLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 910
Db 718 TGNSTNSNLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 777
Qy 911 DFTGAPQNGENKPSKNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 970
Db 778 DFTGAPQNGENKPSKNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 837
Qy 971 PEGNVGSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
Db 838 PEGNVGSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 897
Qy 1031 KKNLSDFIA 1039
Db 898 KKNLSDFIA 906
RESULT 34
US-09-884-465A-369
; Sequence 369, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 369
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-369
Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 551 GDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGEBELTNVYNLLKNSTFN 610
Db 145 GDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGEBELTNVYNLLKNSTFN 204
Qy 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGNIANFE 670
Db 205 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGNIANFE 264
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 730
Db 265 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 324

Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 384
Qy 791 LDQSTYIIVEVPILEKENQTKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLS 850
Db 385 LDQSTYIIVEVPILEKENQTKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLS 444
Qy 851 TGNSTNSNLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 910
Db 445 TGNSTNSNLTLEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 504
Qy 911 DFTGAPQNGENKPSKNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 970
Db 505 DFTGAPQNGENKPSKNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 564
Qy 971 PEGNVGSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
Db 565 PEGNVGSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
Qy 1031 KKNLSDFIA 1039
Db 625 KKNLSDFIA 633
RESULT 35
US-09-884-465A-370
; Sequence 370, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 370
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-370
Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 551 GDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGEBELTNVYNLLKNSTFN 610
Db 418 GDHHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGEBELTNVYNLLKNSTFN 477
Qy 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGNIANFE 670
Db 478 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGNIANFE 537
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 730
Db 538 LDOPYLPQGTFTKTYTIASKDYPEVSDGTFTVPTSLAYKWSOTIFYPPHAGDTYLRVNPQ 597
Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 598 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 657

Qy	791	LDNQSYIIVEVPILEKENOTDKPSILPQFKRNKAQENSKLOEKVEEPKTSSEKVEKLSL	850
Db	658	LDNQSYIIVEVPILEKENOTDKPSILPQFKRNKAQENSKLOEKVEEPKTSSEKVEKLSL	717
Qy	851	TGNSTNSLTLEEVPVDVPQVEKAFASYGKMLNVLNMDGTIELYLPSEGVKKNMA	910
Db	718	TGNSTNSLTLEEVPVDVPQVEKAFASYGKMLNVLNMDGTIELYLPSEGVKKNMA	777
Qy	911	DFTGEAPOGNGENKPSENGKVGSTGVTEKOPTENKPADSLPEAPNEKPKVPKNSTNCGMLN	970
Db	778	DFTGEAPOGNGENKPSENGKVGSTGVTEKOPTENKPADSLPEAPNEKPKVPKNSTNCGMLN	837
Qy	971	PEGNVSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVFNMDGTIELRLPSGEVI	1030
Db	838	PEGNVSDPMLDPALEAPAVDPVQVEKLEKFTASYGLGLDSVFNMDGTIELRLPSGEVI	897
Qy	1031	KKNLSDFIA	1039
Db	898	KKNLSDFIA	906

```

RESULT 36
US-09-884-465A-373
; Sequence 373, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 373
; LENGTH: 906
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-373

```

```

Db      445  TGNSTNSTLEEVPTVPQVEKVKAFPSYGNKGLNVLFNMDGTIELPLPSGEVIRKMA 504
      911  DFTGEAPQNGENKPSGKSVGTGTVEHQPTENKPADSLPEAPNEKPKVKPENSTDNGLN 970
      505  DFTGEAPQNGENKPSGKSVGTGTVEHQPTENKPADSLPEAPNEKPKVKPENSTDNGLN 564
      971  PEGNVGSDPMLDPALEEAAPVDPVQEKLEKFTASYGLGLGDSVIFNMDGTIELRLPSGEVI 1030
      565  PEGNVGSDPMLDPALEEAAPVDPVQEKLEKFTASYGLGLGDSVIFNMDGTIELRLPSGEVI 624

      QY      1031  KKNLSDFIA 1039
      Db      625  KKNLSDFIA 633

RESULT 37
      US-09-884-465A-356
      ; Sequence 356, Application US/09884465A
      ; Publication No. US2003007293A1
      ; GENERAL INFORMATION:
      ; APPLICANT: Shire Biochem, Inc.
      ; APPLICANT: Hamel, Josée
      ; APPLICANT: Brodeur, Bernard
      ; APPLICANT: Martin, Denis
      ; APPLICANT: Charland, Nathalie
      ; APPLICANT: Ouellet, Catherine
      ; TITLE OF INVENTION: Streptococcus Antigens
      ; FILE REFERENCE: 055190-0044
      ; CURRENT APPLICATION NUMBER: US/09/884,465A
      ; CURRENT FILING DATE: 2001-06-20
      ; PRIOR APPLICATION NUMBER: 60/212,683
      ; PRIOR FILING DATE: 2000-06-20
      ; NUMBER OF SEQ ID NOS: 384
      ; SOFTWARE: PatentIn version 3.1
      ; SEQ ID NO 356
      ; LENGTH: 569
      ; TYPE: PRI
      ; ORGANISM: Artificial Sequence
      ; FEATURE:
      ; OTHER INFORMATION: Unknown Organism
      US-09-884-465A-356

```

	Query Match	46.8%;	Score 486;	DB 9;	Length 569;	
	Best Local Similarity	100.0%;	Pred. No. 0;			
	Matches 486;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps	
QY	554	HHADPIDEHKPVGIGHSHSNYLFPPEGGVAKKGNKYVTGEELTNVVNLLKNSTFNQN	613			
Dd	84	HHADPTDEHKPVGIGHSHSNYLFPPEGVAKKGNKYVTGEEITNVYNLLKNSTFNQN	143			
QY	614	FTLANGCKRVSSFPELEKKLGINMLVKLTPDGKVLEKSGVFEGVGNIANFELDQ	673			
Dd	144	FTLANGCKRVSSFPELEKKLGINMLVKLTPDGKVLEKSGVFEGVGNIANFELDQ	203			
QY	674	PYPGGTFFKYTTASKDYPEVSVDGTTFVPTSLAYKMASQTIFYPFHAGDTYLRVPQFAV	733			
Dd	204	PYPGGTFFKYTTASKDYPEVSVDGTTFVPTSLAYKMASQTIFYPFHAGDTYLRVPQFAV	263			
QY	734	PKCTDALVRVDFEFGHNAYLENNYKVEIKPIPKLNCGTTRTAGNKLPVTFMANAYLDN	793			
Dd	264	PKCTDALVRVDEFHGHNAYLENNYKVEIKPIPKLNCGTTRTAGNKIPVTFMANAYLDN	323			
QY	794	QSIIYIYEIVPILEKENOTDRPSILPOFKRNAQENSKLDEKVEEKPLETSIGN	853			
Dd	324	QSIIYIYEIVPILEKENOTDRPSILOPKRNAQENSKLDEKVEEKPLETSIGN	383			
QY	854	STNSNTLEEVPDVDPQEKAFAESYGKLNVLFNMDDGIETLYLSPGSGEVIKKMAFFT	913			
Dd	384	STNSNTLEEVPDVDPQEKAFAESYGKLVNFNMDDGIETLYLSPGSGEVIKKMAFFT	443			
QY	914	GEAPQGNGENKPSENGKSVTGTVENQPTENPADSLPEAPNEKPVKPENSTDNGLMNAPEG	973			

Db 444 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 503

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 504 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 563

Qy 1034 LSDFIA 1039

Db 564 LSDFIA 569

RESULT 38

US-09-884-465A-353

Sequence 353, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

PRIOR FILING DATE: 2001-06-20

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 384

SOFTWARE: PatentIn version 3.1

SEQ ID NO 353

LENGTH: 633

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

US-09-884-465A-353

Query Match 46.8%; Score 486; DB 9; Length 633;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 554 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKKNSTNNQN 613

Db 148 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKKNSTNNQN 207

Qy 614 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDQ 673

Db 208 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDQ 267

Qy 674 PYLPQGTFTKTYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733

Db 268 PYLPQGTFTKTYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327

Qy 734 PKGTDALVRVDFEFGHGNAYLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 793

Db 328 PKGTDALVRVDFEFGHGNAYLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 387

Qy 794 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853

Db 388 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447

Qy 854 STSNSTLEEVPVTDVQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMADFT 913

Db 448 STSNSTLEEVPVTDVQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMADFT 507

Qy 914 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 1033

Db 508 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 567

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 568 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 627

Qy 1034 LSDFIA 1039

Db 628 LSDFIA 633

RESULT 39

US-09-884-465A-355

Sequence 355, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

PRIOR FILING DATE: 2001-06-20

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 384

SOFTWARE: PatentIn version 3.1

SEQ ID NO 355

LENGTH: 633

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

US-09-884-465A-355

Query Match 46.8%; Score 486; DB 9; Length 633;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 554 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKKNSTNNQN 613

Db 148 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKKNSTNNQN 207

Qy 614 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDQ 673

Db 208 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDQ 267

Qy 674 PYLPQGTFTKTYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733

Db 268 PYLPQGTFTKTYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327

Qy 734 PKGTDALVRVDFEFGHGNAYLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 793

Db 328 PKGTDALVRVDFEFGHGNAYLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 387

Qy 794 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853

Db 388 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447

Qy 854 STSNSTLEEVPVTDVQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMADFT 913

Db 448 STSNSTLEEVPVTDVQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMADFT 507

Qy 914 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 973

Db 508 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 567

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 568 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 627

Qy 1034 LSDFIA 1039

Db 628 LSDFIA 633

RESULT 40
 US-09-884-465A-363
 ; Sequence 363, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR FILING DATE: 2001-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 363
 ; LENGTH: 633
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 US-09-884-465A-363

Query Match 46.8%; Score 486; DB 9; Length 633;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	554	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNN	613
DB	148	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNN	207
QY	614	FTLANGKRVSFSPPELEKKGINMLVLIITPDGKVLKVSCKVFGVGVGNIANFELDQ	673
DB	208	FTLANGKRVSFSPPELEKKGINMLVLIITPDGKVLKVSCKVFGVGVGNIANFELDQ	267
QY	674	PYLPGGTFYTIASKDYPEVSDGTFTVPTSLAYKVASQTIFFPHAGDTYLRVNPQFAV	733
DB	268	PYLPGGTFYTIASKDYPEVSDGTFTVPTSLAYKVASQTIFFPHAGDTYLRVNPQFAV	327
QY	734	PKGTDLVRVDFEFGHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLON	793
DB	328	PKGTDLVRVDFEFGHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLON	387
QY	794	QSTYIVVEVPVILEKENQDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGN	853
DB	388	QSTYIVVEVPVILEKENQDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGN	447
QY	854	STNSSTLEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMAFT	913
DB	448	STNSSTLEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMAFT	507
QY	914	GEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNMGLNPEG	973
DB	508	GEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNMGLNPEG	567
QY	974	NGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPLSPSEVIKKN	1033
DB	568	NGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPLSPSEVIKKN	627
QY	1034	LSDFIA 1039	
DB	628	LSDFIA 633	

RESULT 41
 US-09-884-465A-364
 ; Sequence 364, Application US/09884465A
 ; Publication No. US20030077293A1

; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR FILING DATE: 2001-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 364
 ; LENGTH: 633
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 US-09-884-465A-364

Query Match 46.8%; Score 486; DB 9; Length 633;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	554	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNN	613
DB	148	HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNN	207
QY	614	FTLANGKRVSFSPPELEKKGINMLVLIITPDGKVLKVSCKVFGVGVGNIANFELDQ	673
DB	208	FTLANGKRVSFSPPELEKKGINMLVLIITPDGKVLKVSCKVFGVGVGNIANFELDQ	267
QY	674	PYLPGGTFYTIASKDYPEVSDGTFTVPTSLAYKVASQTIFFPHAGDTYLRVNPQFAV	733
DB	268	PYLPGGTFYTIASKDYPEVSDGTFTVPTSLAYKVASQTIFFPHAGDTYLRVNPQFAV	327
QY	734	PKGTDLVRVDFEFGHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLON	793
DB	328	PKGTDLVRVDFEFGHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLON	387
QY	794	QSTYIVVEVPVILEKENQDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGN	853
DB	388	QSTYIVVEVPVILEKENQDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGN	447
QY	854	STNSSTLEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMAFT	913
DB	448	STNSSTLEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMAFT	507
QY	914	GEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNMGLNPEG	973
DB	508	GEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNMGLNPEG	567
QY	974	NGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPLSPSEVIKKN	1033
DB	568	NGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPLSPSEVIKKN	627
QY	1034	LSDFIA 1039	
DB	628	LSDFIA 633	

RESULT 42
 US-09-884-465A-365
 ; Sequence 365, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 365
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-365

Query Match 46.8%; Score 486; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGKVKVTGEEELTNVNNLLKKNSTFNNQ 613
DB 148 HHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGKVKVTGEEELTNVNNLLKKNSTFNNQ 207
QY 614 FTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKVLKSVKVGEGVGNIANFELDQ 673
DB 208 FTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKVLKSVKVGEGVGNIANFELDQ 267
QY 674 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733
DB 268 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327
QY 734 PKGTDALVRVDFEFGHNAVLENNYKVGKILPIPKLNOGTTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDALVRVDFEFGHNAVLENNYKVGKILPIPKLNOGTTAGNKIPVTFMANAYLDN 387
QY 794 QSTIVIEVPILEKENQTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTIVIEVPILEKENQTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLSLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 913
DB 448 STNSNLSLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 507
QY 914 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEG 973
DB 508 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEG 567
QY 974 NVGSDPMLDPALEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 1033
DB 568 NVGSDPMLDPALEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 627

QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 43
US-09-884-465A-371
Sequence 371, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A

CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 371
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-371

Query Match 46.8%; Score 486; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGKVKVTGEEELTNVNNLLKKNSTFNNQ 613
DB 148 HHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGKVKVTGEEELTNVNNLLKKNSTFNNQ 207
QY 614 FTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKVLKSVKVGEGVGNIANFELDQ 673
DB 208 FTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKVLKSVKVGEGVGNIANFELDQ 267
QY 674 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733
DB 268 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327
QY 734 PKGTDALVRVDFEFGHNAVLENNYKVGKILPIPKLNOGTTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDALVRVDFEFGHNAVLENNYKVGKILPIPKLNOGTTAGNKIPVTFMANAYLDN 387
QY 794 QSTIVIEVPILEKENQTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTIVIEVPILEKENQTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLSLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 913
DB 448 STNSNLSLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 507
QY 914 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEG 973
DB 508 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEG 567
QY 974 NVGSDPMLDPALEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 1033
DB 568 NVGSDPMLDPALEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 627

QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 44
US-09-884-465A-372
Sequence 372, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1

SEQ ID NO 372
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match 46.8%; Score 486; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYVTGEEITNVVNLKNSFNQNT 613
Db* 421 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYVTGEEITNVVNLKNSFNQNT 480
QY 614 FTLANGOKRVVSFFPELEKLGINMLVKLIITPDGKVKLEKSGKVGEGVGNIANFELDQ 673
Db 481 FTLANGOKRVVSFFPELEKLGINMLVKLIITPDGKVKLEKSGKVGEGVGNIANFELDQ 540
QY 674 PYLPGOTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAV 733
Db 541 PYLPGOTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAV 600
QY 734 PKGTALVRVDFDEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 793
Db 601 PKGTALVRVDFDEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 660
QY 794 QSTYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGN 853
Db 661 QSTYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGN 720
QY 854 STSNSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFT 913
Db 721 STSNSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFT 780
QY 914 GEAPQNGENKPSSENGKSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEG 973
Db 781 GEAPQNGENKPSSENGKSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEG 840
QY 974 NVGSDPMLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVKKN 1033
Db 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVKKN 900
QY 1034 LSDFIA 1039
Db 901 LSDFIA 906

RESULT 45
US-09-884-465A-256
Sequence 256, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 053190-0044
CURRENT APPLICATION NUMBER: US/09/884.465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 256
LENGTH: 569
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Unknown Sequence
US-09-884-465A-256

Query Match 46.8%; Score 484; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 556 ADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYVTGEEITNVVNLKNSFNQNT 615
Db 86 ADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNKYVTGEEITNVVNLKNSFNQNT 145
QY 616 LANGOKRVVSFFPELEKLGINMLVKLIITPDGKVKLEKSGKVGEGVGNIANFELDQPY 675
Db 146 LANGOKRVVSFFPELEKLGINMLVKLIITPDGKVKLEKSGKVGEGVGNIANFELDQPY 205
QY 676 LPGOTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAV 735
Db 206 LPGOTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAV 265
QY 736 GTDALVRVDFDEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOS 795
Db 266 GTDALVRVDFDEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOS 325
QY 796 TYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNST 855
Db 326 TYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNST 385
QY 856 SNSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGE 915
Db 386 SNSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGE 445
QY 916 APOGNGENKPSSENGKSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEGNV 975
Db 446 APOGNGENKPSSENGKSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEGNV 505
QY 976 GSDPMLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVKKNLS 1035
Db 506 GSDPMLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVKKNLS 565
QY 1036 DFTA 1039
Db 566 DFTA 569

Search completed: May 13, 2003, 14:02:58
Job time : 54 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein. - protein search, using sw model

Run On: May 13, 2003, 13:45:42 ; Search time 22.5622 Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKIEKIKAGIMKQYG.....IELRLPSGEVKKNLSDFIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2746	100.0	528	AA12719	Streptococcus pneu
2	2746	100.0	528	AAU84024	Truncated variant
3	2746	100.0	568	AA12731	Streptococcus pneu
4	2746	100.0	568	AAU83999	Truncated variant
5	2746	100.0	626	AAU84018	Truncated variant
6	2746	100.0	632	AAU83998	Truncated variant
7	2746	100.0	632	AAU84002	Truncated variant
8	2746	100.0	632	AAU84003	Truncated variant
9	2746	100.0	632	AAU84013	Truncated variant
10	2746	100.0	632	AAU84014	Truncated variant

11	2746	100.0	644	23	AAU83996	Truncated variant
12	2746	100.0	807	23	AAU83997	Truncated variant
13	2746	100.0	895	23	AAU84070	S. pneumoniae deri
14	2746	100.0	896	23	AAU84080	S. pneumoniae deri
15	2746	100.0	901	23	AAU84068	S. pneumoniae deri
16	2746	100.0	901	23	AAU84072	S. pneumoniae deri
17	2746	100.0	901	23	AAU84074	S. pneumoniae deri
18	2746	100.0	902	23	AAU84076	S. pneumoniae deri
19	2746	100.0	902	23	AAU84078	S. pneumoniae deri
20	2746	100.0	913	23	AAU84059	S. pneumoniae deri
21	2746	100.0	999	23	AAU84051	S. pneumoniae deri
22	2746	100.0	999	23	AAU84052	S. pneumoniae deri
23	2746	100.0	1019	21	AAU84052	Streptococcus pneu
24	2746	100.0	1019	21	AAU84051	Streptococcus pneu
25	2746	100.0	1019	23	AAU84021	Truncated variant
26	2746	100.0	1039	21	AAU84051	Streptococcus pneu
27	2746	100.0	1039	23	AAU84057	Streptococcus pneu
28	2746	100.0	1057	21	AAU84097	Streptococcus pneu
29	2746	100.0	1058	23	AAU84057	S. pneumoniae deri
30	2746	100.0	1126	23	AAU84058	S. pneumoniae deri
31	2746	100.0	1139	23	AAU84055	S. pneumoniae deri
32	2746	100.0	1152	23	AAU84054	S. pneumoniae deri
33	2746	100.0	1238	23	AAU84056	S. pneumoniae deri
34	2746	100.0	1365	23	AAU84057	S. pneumoniae deri
35	2746	100.0	1378	23	AAU84053	S. pneumoniae deri
36	2740	99.8	1019	21	AAU84053	S. pneumoniae deri
37	2737	99.7	832	23	AAU84004	Truncated variant
38	2737	99.7	832	23	AAU84006	Truncated variant
39	2737	99.7	632	23	AAU84009	Truncated variant
40	2737	99.7	632	23	AAU84012	Truncated variant
41	2737	99.7	901	23	AAU84066	S. pneumoniae deri
42	2737	99.7	901	23	AAU84067	S. pneumoniae deri
43	2737	99.7	907	23	AAU84060	S. pneumoniae deri
44	2737	99.7	907	23	AAU84061	S. pneumoniae deri
45	2737	99.7	907	23	AAU84064	S. pneumoniae deri

ALIGNMENTS

RESULT 1
AA12719
ID AA12719 standard; Protein: 528 AA.

XX AA12719;

XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX WO200039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.

XX N-PSDB; AAA65734.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media; bacteraemia and/or pneumonia

XX PS Claim 18; Fig 10; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents the

CC S. pneumoniae BVH-3B protein antigen.

XX SQ Sequence 528 AA;

Query Match 100.0%; Score 2746; DB 21; Length 528;

Best Local Similarity 100.0%; Pred. No. 2.4e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHDADPIDEHKPVGICHSH 60

DB 1 MKDLKKIEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHDADPIDEHKPVGICHSH 60

QY 61 SNYELFKPEGVAKKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120

DB 61 SNYELFKPEGVAKKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120

QY 121 EKKLGIMLVKLIITPDGKLVKSVKGVGEGVNTANFELDQPYLPQGTFRYTIASKDYP 180

DB 121 EKKLGIMLVKLIITPDGKLVKSVKGVGEGVNTANFELDQPYLPQGTFRYTIASKDYP 180

QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240

DB 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240

QY 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILKEKQTD 300

DB 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILKEKQTD 300

QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKELSETGNSTNSLLEEVPTVDPVQE 360

DB 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKELSETGNSTNSLLEEVPTVDPVQE 360

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQNGENKPSGENKV 420

DB 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQNGENKPSGENKV 420

QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGNVGSDDPMLDPALEAPAV 480

DB 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGNVGSDDPMLDPALEAPAV 480

QY 481 DPVQEKLEKFTASYGLDVSIFNMDGTIELRPSGEVYIKKNLSDFIA 528

DB 481 DPVQEKLEKFTASYGLDVSIFNMDGTIELRPSGEVYIKKNLSDFIA 528

RESULT 2

AAU84024

ID AAU84024 standard; Peptide: 528 AA.

XX AC AAU84024;

XX 08-MAY-2002 (first entry)

XX Truncated variant of S. pneumoniae BVH-3, BVH-3B.

DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; muten.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

XX PD 27-DEC-2001.

XX 19-JUN-2001: 2001WO-CA00908.

XX 20-JUN-2000: 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI: 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia

XX Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or S. saprophyticus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX SQ Sequence 528 AA;

Query Match 100.0%; Score 2746; DB 23; Length 528;

Best Local Similarity 100.0%; Pred. No. 2.4e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHDADPIDEHKPVGICHSH 60

DB 1 MKDLKKIEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHDADPIDEHKPVGICHSH 60

QY 61 SNYELFKPEGVAKKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120

DB 61 SNYELFKPEGVAKKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120

QY 121 EKKLGIMLVKLIITPDGKLVKSVKGVGEGVNTANFELDQPYLPQGTFRYTIASKDYP 180

DB 121 EKKLGIMLVKLIITPDGKLVKSVKGVGEGVNTANFELDQPYLPQGTFRYTIASKDYP 180

QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240

DB 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240

QY 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILKEKQTD 300

DB 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILKEKQTD 300

QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKELSETGNSTNSLLEEVPTVDPVQE 360

DB 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKELSETGNSTNSLLEEVPTVDPVQE 360

```
QY 361 KVAFAESYGMKLENVLFNMDGTIELPLSGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
Db 361 KVAFAESYGMKLENVLFNMDGTIELPLSGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPMLDPALEAPAV 480
|||||
Db 421 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPMLDPALEAPAV 480
|||||
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||
Db 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||

RESULT 3
AAB12731
ID AAB12731 standard; Protein: 568 AA.
XX
AC AAB12731;
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur Br, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
PS Claim 18; Fig 29; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW1 protein antigen.
XX
SQ Sequence 568 AA:
Query Match 100.0%; Score 2746; DB 21; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLDKTEETAGIMKQYGVKRESIVVNKEKNAIIPHGDDHADPIDERKPVGIGHSH 60
|||||
Db 41 MKDLDKTEETAGIMKQYGVKRESIVVNKEKNAIIPHGDDHADPIDERKPVGIGHSH 100
|||||
QY 61 SNYELFKPEEGVAKGKNVYTGELTNVNLKNSFNQNFNLANGQKRVSFPPPEL 120
|||||
Db 101 SNYELFKPEEGVAKGKNVYTGELTNVNLKNSFNQNFNLANGQKRVSFPPPEL 160
|||||
QY 121 EKKLGINMLVKLITPDGKVLKESKVGEGVGNIANFELDQVLPQGTFKYTTASKDYP 180
|||||
```

```
Db 161 EKKLGINMLVKLITPDGKVLKESKVGEGVGNIANFELDQVLPQGTFKYTTASKDYP 220
|||||
QY 181 EYSDGTFTVPTSLAYKNASQTIIFYPHAGDTYLRVNPQFAVPKGTDALYRVFDEFHGNA 240
|||||
Db 221 EYSDGTFTVPTSLAYKNASQTIIFYPHAGDTYLRVNPQFAVPKGTDALYRVFDEFHGNA 280
|||||
QY 241 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 300
|||||
Db 281 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 340
|||||
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSTNSLTSEVPVDPVQE 360
|||||
Db 341 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSTNSLTSEVPVDPVQE 400
|||||
QY 361 KVAFAESYGMKLENVLFNMDGTIELPLSGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
Db 401 KVAFAESYGMKLENVLFNMDGTIELPLSGEVYKKNMADFTGEAPQNGENKPSSENGKV 460
|||||
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPMLDPALEAPAV 480
|||||
Db 461 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPMLDPALEAPAV 520
|||||
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||
Db 521 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 568
|||||

RESULT 4
AAU83999
ID AAU83999 standard; Peptide; 568 AA.
XX
AC AAU83999;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW1.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; Streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
DE New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
CC Claim 1; Page 1; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
```

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX SQ Sequence 568 AA;
 Query Match 100.0%; Score 2746; DB 23; Length 568;
 Best Local Similarity 100.0%; Pred. No. 2.6e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPHGDHHDADPIDEHKKPVGIGHSH 60
 DB 41 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPHGDHHDADPIDEHKKPVGIGHSH 100
 QY 61 SNYELFKPEEGVAKKGNKYVTGEELTNVYLLKSTFNQNTLANGQKRVSFSPPEL 120
 DB 101 SNYELFKPEEGVAKKGNKYVTGEELTNVYLLKSTFNQNTLANGQKRVSFSPPEL 160
 QY 121 EKKLGINMLVKLIIPDCKVLEKSGVFGEGVGNIAFELDPQYLPQGFYTIASKDYP 180
 DB 161 EKKLGINMLVKLIIPDCKVLEKSGVFGEGVGNIAFELDPQYLPQGFYTIASKDYP 220
 QY 181 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGNA 240
 DB 221 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGNA 280
 QY 241 YLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 300
 DB 281 YLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 340
 QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGNSNSTLEEYPTVDPVQVE 360
 DB 341 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGNSNSTLEEYPTVDPVQVE 400
 QY 361 KVAFKESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSGENKV 420
 DB 401 KVAFKESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSGENKV 460
 QY 421 STGIVENQPTENKPADSLPAPNEKPKVKNSTNDNMLNPEGNVGSDDMLDPALEAPAV 480
 DB 461 STGIVENQPTENKPADSLPAPNEKPKVKNSTNDNMLNPEGNVGSDDMLDPALEAPAV 520
 QY 481 DPVQEKLEKTASYGLDGVIFNMDGTIELRPSGEVIAKKNLSDPIA 528
 DB 521 DPVQEKLEKTASYGLDGVIFNMDGTIELRPSGEVIAKKNLSDPIA 568

RESULT 5
 AAU84018
 ID AAU84018 standard; Peptide; 626 AA.

XX AC AAU84018;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, NEW105.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX OS pneumonia; streptococcal bacterial infection; mutant; mutuin.

XX OS Streptococcus pneumoniae.

OS Synthetic.

XX PN WO200198334-A2.
 XX PD 27-DEC-2001.
 XX PF 19-JUN-2001: 2001WO-CA00908.
 XX PR 20-JUN-2000: 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.
 XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI: 2002-122272/16.
 XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX PS Claim 1; Page -: 113pp; English.
 XX CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX SQ Sequence 626 AA;

Query Match 100.0%; Score 2746; DB 23; Length 626;
 Best Local Similarity 100.0%; Pred. No. 3e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPHGDHHDADPIDEHKKPVGIGHSH 60
 DB 99 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPHGDHHDADPIDEHKKPVGIGHSH 158
 QY 61 SNYELFKPEEGVAKKGNKYVTGEELTNVYLLKSTFNQNTLANGQKRVSFSPPEL 120
 DB 159 SNYELFKPEEGVAKKGNKYVTGEELTNVYLLKSTFNQNTLANGQKRVSFSPPEL 218
 QY 121 EKKLGINMLVKLIIPDCKVLEKSGVFGEGVGNIAFELDPQYLPQGFYTIASKDYP 180
 DB 219 EKKLGINMLVKLIIPDCKVLEKSGVFGEGVGNIAFELDPQYLPQGFYTIASKDYP 278
 QY 181 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGNA 240
 DB 279 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDLVRVDFEFGNA 338
 QY 241 YLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 300
 DB 339 YLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 398
 QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKLSGTGNSNSTLEEYPTVDPVQVE 360
 |||||

Db 399 KPSILPQPKRKAQENSKLDEKVEEPKTSKVEKEKLSGTGNTSTNLEEVTVDPVQE 458
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSNGKV 420
 Db 459 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSNGKV 518
 QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNMGMLNPEGNVGSDPMLDPALEAPAV 480
 Db 519 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNMGMLNPEGNVGSDPMLDPALEAPAV 578
 QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
 Db 579 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 626

RESULT 6

AAU83998

ID AAU83998 standard; Peptide: 632 AA.

XX AC AAU83998;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of *S. pneumoniae* BVH-11, NEW40.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.
 XX OS Synthetic.

XX PN WO200198334-A2.
 XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.
 XX PR 20-JUN-2000; 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 XX meningitis, and bacteraemia

PS Example 1; Page 1; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for *S. pneumoniae* infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX SO Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;
 Best Local Similarity 100.0%; Pred. No. 3e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLDDKIEEKIAGIMKQYGVKRESIVVNKEKNAILIYPHGDHHDADIDEHKPVGIGHSH 60
 Db 105 MKDLDDKIEEKIAGIMKQYGVKRESIVVNKEKNAILIYPHGDHHDADIDEHKPVGIGHSH 164
 QY 61 SNYELFKPEEGVAKKGNKYVTGEELTNVNLLKNSFNQNETLANGQKRVSFSPPEL 120
 Db 165 SNYELFKPEEGVAKKGNKYVTGEELTNVNLLKNSFNQNETLANGQKRVSFSPPEL 224
 QY 121 EKKIGINMLVKLIIPDGKVLKESGVKGVGEGVGNIAFELDQDYPVLPQOTFKYTIASKDYP 180
 Db 225 EKKIGINMLVKLIIPDGKVLKESGVKGVGEGVGNIAFELDQDYPVLPQOTFKYTIASKDYP 284
 QY 181 EVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPKGTIDALVRVDFEFGNA 240
 Db 285 EVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPKGTIDALVRVDFEFGNA 344
 QY 241 YLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTD 300
 Db 345 YLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTD 404
 QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNTSTNLEEVTVDPVQE 360
 Db 405 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNTSTNLEEVTVDPVQE 464
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSNGKV 420
 Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSNGKV 524
 QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNMGMLNPEGNVGSDPMLDPALEAPAV 480
 Db 525 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNMGMLNPEGNVGSDPMLDPALEAPAV 584
 QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
 Db 585 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 632

RESULT 7

AAU84002

ID AAU84002 standard; Peptide: 632 AA.

XX AC AAU84002;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of *S. pneumoniae* BVH-3, NEW49.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.
 XX OS Synthetic.

XX PN WO200198334-A2.
 XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.
 XX PR 20-JUN-2000; 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.

DR WPI: 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

XX meningitis, and bacteraemia

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 3e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKDKIEEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDERKPVGIGHSH 60

DB 105 MKDLKDKIEEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDERKPVGIGHSH 164

QY 61 SNYELFPEGVAKKGNKYVTGEELTNVNLKNSTFNQNTLANGOKRVSFPPEL 120

DB 165 SNYELFPEGVAKKGNKYVTGEELTNVNLKNSTFNQNTLANGOKRVSFPPEL 224

QY 121 EKLGINMLVLIPTDCKVLEKYSKGVFEGVGNIAFNFELDPYLPQTFKVTIASKDYP 180

DB 225 EKLGINMLVLIPTDCKVLEKYSKGVFEGVGNIAFNFELDPYLPQTFKVTIASKDYP 284

QY 181 EYSYDGNFTVPTSLAYKMASOTIFYPFHAGDTVLRVNPQFAYPKCTDALRVDFEFGNA 240

DB 285 EYSYDGNFTVPTSLAYKMASOTIFYPFHAGDTVLRVNPQFAYPKCTDALRVDFEFGNA 344

QY 241 YLENNYKVGEEKLPKPLNGQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENDTD 300

DB 345 YLENNYKVGEEKLPKPLNGQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENDTD 404

QY 301 KPSILPQFKRNKAENSKLDEKVEEPTSKVEKLSGTGNTSTNLTLEETVTPDVPQE 360

DB 405 KPSILPQFKRNKAENSKLDEKVEEPTSKVEKLSGTGNTSTNLTLEETVTPDVPQE 464

QY 361 KVAFASVGMKLENVLFNDGTIELYLPSPGEVIKKNMADFTCEAPQNGENKPSNGKV 420

DB 465 KVAFASVGMKLENVLFNDGTIELYLPSPGEVIKKNMADFTCEAPQNGENKPSNGKV 524

QY 421 STGVENQPTENKPADSLPEAPKPKVPKPNSTNDGNLNPENGVGSDPMDLPALEAPAV 480

DB 525 STGVENQPTENKPADSLPEAPKPKVPKPNSTNDGNLNPENGVGSDPMDLPALEAPAV 584

QY 481 DPVQEKLEKFTASVGLDVSIFNMDSITIELRLPSGEVIKKNLSDFIA 528

DB 585 DPVQEKLEKFTASVGLDVSIFNMDSITIELRLPSGEVIKKNLSDFIA 632

RESULT 8

AAU84003

ID AAU84003 standard; Peptide: 632 AA.

XX AC AAU84003;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, NEW50.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI: 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia

XX Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 3e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKDKIEEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDERKPVGIGHSH 60

DB 105 MKDLKDKIEEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDERKPVGIGHSH 164

QY 61 SNYELFPEGVAKKGNKYVTGEELTNVNLKNSTFNQNTLANGOKRVSFPPEL 120

DB 165 SNYELFPEGVAKKGNKYVTGEELTNVNLKNSTFNQNTLANGOKRVSFPPEL 224

QY 121 EKLGINMLVLIPTDCKVLEKYSKGVFEGVGNIAFNFELDPYLPQTFKVTIASKDYP 180

DB 225 EKLGINMLVLIPTDCKVLEKYSKGVFEGVGNIAFNFELDPYLPQTFKVTIASKDYP 284

QY 181 EYSYDGNFTVPTSLAYKMASOTIFYPFHAGDTVLRVNPQFAYPKCTDALRVDFEFGNA 240

DB 285 EYSYDGNFTVPTSLAYKMASOTIFYPFHAGDTVLRVNPQFAYPKCTDALRVDFEFGNA 344

QY 241 YLENNYKVGEEKLPKPLNGQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENDTD 300

DB 345 YLENNYKVGEEKLPKPLNGQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENDTD 404

QY 301 KPSILPQFKRNKAENSKLDEKVEEPTSKVEKLSGTGNTSTNLTLEETVTPDVPQE 360

DB 405 KPSILPQFKRNKAENSKLDEKVEEPTSKVEKLSGTGNTSTNLTLEETVTPDVPQE 464

QY 361 KVAFASVGMKLENVLFNDGTIELYLPSPGEVIKKNMADFTCEAPQNGENKPSNGKV 420

DB 465 KVAFASVGMKLENVLFNDGTIELYLPSPGEVIKKNMADFTCEAPQNGENKPSNGKV 524

QY 421 STGVENQPTENKPADSLPEAPKPKVPKPNSTNDGNLNPENGVGSDPMDLPALEAPAV 480

DB 525 STGVENQPTENKPADSLPEAPKPKVPKPNSTNDGNLNPENGVGSDPMDLPALEAPAV 584

QY 481 DPVQEKLEKFTASVGLDVSIFNMDSITIELRLPSGEVIKKNLSDFIA 528

DB 585 DPVQEKLEKFTASVGLDVSIFNMDSITIELRLPSGEVIKKNLSDFIA 632

|||||
Db 165 SNYELFKPEGVAKKGVYTGELTNVNLKSTFNQNF LANGQRVSFSPPEL 224
QY 121 EKKLGINMLVKLITPDGKVLKSKGVFGVGNIANFELDPYLPQGTFKYTIASKDYP 180
Db 225 EKKLGINMLVKLITPDGKVLKSKGVFGVGNIANFELDPYLPQGTFKYTIASKDYP 284
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNA 240
Db 285 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNA 344
QY 241 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVVEVPILEKENQTD 300
Db 345 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVVEVPILEKENQTD 404
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLDEEPTVDVQVE 360
Db 405 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLDEEPTVDVQVE 464
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTFKYTIASKDYP 420
Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTFKYTIASKDYP 524
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 480
Db 525 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 584
QY 481 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 528
Db 585 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 632

RESULT 9

AAU84013
ID AAU84013 standard; Peptide: 632 AA.
XX AC AAU84013;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW63.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia: streptococcal bacterial infection; mutant; mutain.
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO200198334-A2.
XX XX 27-DEC-2001.
XX XX 19-JUN-2001; 2001WO-CA00908.
XX XX 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX XX
XX XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia
XX XX
XX PS Claim 1; Page -: 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX SQ Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;
Best local Similarity 100.0%; Pred. No. 3e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
Db 105 MKDLKKIEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 164
QY 61 SNYELFKPEGVAKKGVYTGELTNVNLKSTFNQNF LANGQRVSFSPPEL 120
Db 165 SNYELFKPEGVAKKGVYTGELTNVNLKSTFNQNF LANGQRVSFSPPEL 224
QY 121 EKKLGINMLVKLITPDGKVLKSKGVFGVGNIANFELDPYLPQGTFKYTIASKDYP 180
Db 225 EKKLGINMLVKLITPDGKVLKSKGVFGVGNIANFELDPYLPQGTFKYTIASKDYP 284
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNA 240
Db 285 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNA 344
QY 241 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVVEVPILEKENQTD 300
Db 345 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVVEVPILEKENQTD 404
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLDEEPTVDVQVE 360
Db 405 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLDEEPTVDVQVE 464
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTFKYTIASKDYP 420
Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTFKYTIASKDYP 524
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 480
Db 525 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 584
QY 481 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 528
Db 585 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFTA 632

RESULT 10

AAU84014
ID AAU84014 standard; Peptide: 632 AA.
XX AC AAU84014;
XX XX
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW64.
XX XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW

CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX
 XX
 SQ Sequence 644 AA;

Query Match 100.0%; Score 2746; DB 23; Length 644;
 Best Local Similarity 100.0%; Pred. No. 3.1e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIYPHGDHHDADPDEHKPVGIGHSH 60
 |||||
 DB 117 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIYPHGDHHDADPDEHKPVGIGHSH 176
 QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFTNNQNTLANGQKRVSFSEPEL 120
 |||||
 DB 177 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFTNNQNTLANGQKRVSFSEPEL 236
 QY 121 EKKLGINMLVLIITPDGKVLKESKVGEGVGNIANFELDQPLPGQTFYTTASKDYP 180
 |||||
 DB 237 EKKLGINMLVLIITPDGKVLKESKVGEGVGNIANFELDQPLPGQTFYTTASKDYP 296
 QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
 |||||
 DB 297 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 356
 QY 241 YLENNYKVGEEKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVEVPILKEKQTD 300
 |||||
 DB 357 YLENNYKVGEEKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVEVPILKEKQTD 416
 QY 301 KPSILPQFKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLSNLTSEEVVPDVPQE 360
 |||||
 DB 417 KPSILPQFKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLSNLTSEEVVPDVPQE 476
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKMADFTGEAPQNGENKPSENGKV 420
 |||||
 DB 477 KVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKMADFTGEAPQNGENKPSENGKV 536
 QY 421 STGTVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEAPAV 480
 |||||
 DB 537 STGTVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEAPAV 596
 QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFTA 528
 |||||
 DB 597 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFTA 644

RESULT 12

AAU83997
 ID AAU83997 standard; Peptide: 807 AA.

XX
 AC AAU83997;

XX
 DT 08-MAY-2002 (first entry)

XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW25.

XX
 KW BVH-3; BVH-11; vaccine; meningitis; Otitis media; bacteraemia;
 XX pneumonia; streptococcal bacterial infection; mutant; mutein.

XX
 OS Streptococcus pneumoniae.

XX
 OS Synthetic.

PN WO200198334-A2.

XX
 PD 27-DEC-2001.

XX
 PF 19-JUN-2001; 2001WO-CA00908.

XX
 PR 20-JUN-2000; 2000US-212683P.

XX

PA (SHIR-) SHIRE BIOCHEM INC.

XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX
 DR WPI: 2002-122272/16.

XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia

XX
 PS Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX
 SQ Sequence 807 AA;

Query Match 100.0%; Score 2746; DB 23; Length 807;
 Best Local Similarity 100.0%; Pred. No. 4.1e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIYPHGDHHDADPDEHKPVGIGHSH 60
 |||||
 DB 280 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIYPHGDHHDADPDEHKPVGIGHSH 339
 QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFTNNQNTLANGQKRVSFSEPEL 120
 |||||
 DB 340 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFTNNQNTLANGQKRVSFSEPEL 399
 QY 121 EKKLGINMLVLIITPDGKVLKESKVGEGVGNIANFELDQPLPGQTFYTTASKDYP 180
 |||||
 DB 400 EKKLGINMLVLIITPDGKVLKESKVGEGVGNIANFELDQPLPGQTFYTTASKDYP 459
 QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
 |||||
 DB 460 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 519
 QY 241 YLENNYKVGEEKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVEVPILKEKQTD 300
 |||||
 DB 520 YLENNYKVGEEKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVEVPILKEKQTD 579
 QY 301 KPSILPQFKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLSNLTSEEVVPDVPQE 360
 |||||
 DB 580 KPSILPQFKRKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLSNLTSEEVVPDVPQE 639
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKMADFTGEAPQNGENKPSENGKV 420
 |||||
 DB 640 KVAKFAESYGMKLENVLFNMDGTIELYLPSEVIKKMADFTGEAPQNGENKPSENGKV 699
 QY 421 STGTVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEAPAV 480
 |||||
 DB 700 STGTVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEAPAV 759

QY	481	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	528
DB	760	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	807
RESULT 13			
AAU84070	ID	AAU84070	standard; Peptide: 895 AA.
XX	AC	AAU84070;	
XX	DT	08-MAY-2002	(first entry)
XX	DE	S. pneumoniae derived chimeric peptide, VP112.	
XX	KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;	
XX	KW	pneumonia; streptococcal bacterial infection; mutant; mutain;	
XX	KW	BVH-11-2.	
XX	OS	Streptococcus pneumoniae.	
XX	OS	Synthetic.	
XX	PN	WO200198334-A2.	
XX	PD	27-DEC-2001.	
XX	PF	19-JUN-2001; 2001WO-CA00908.	
XX	PR	20-JUN-2000; 2000US-212683P.	
XX	PA	(SHIR-) SHIRE BIOCHEM INC.	
XX	PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
XX	DR	WPI; 2002-122272/16.	
XX	PS	Claim 1: Page -; 113pp; English.	
XX	CC	The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.	
XX	CC	Note: This sequence does not appear in the specification but has been created according to information given in the invention.	
XX	CC	Sequence 895 AA;	
XX	CC	Query Match 100.0%; Score 2746; DB 23; Length 895;	
XX	CC	Best Local Similarity 100.0%; Pred. No. 4.7e-174;	
XX	CC	Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

QY	1	MKOLDKKIEBKTAGIMKQYGVKRESIVNKKENALITYPHGDHHADPIDEHKPVGIGHSH	60
DB	368	MKOLDKKIEBKTAGIMKQYGVKRESIVNKKENALITYPHGDHHADPIDEHKPVGIGHSH	427
QY	61	SNVELFKPEEGVAKKEGKNVYTGEELTNVNVNLLKNTFNQNFLLANGOKRVSFSPPEL	120
DB	428	SNVELFKPEEGVAKKEGKNVYTGEELTNVNVNLLKNTFNQNFLLANGOKRVSFSPPEL	487
QY	121	EKKLGINMLVKLITPDGKVKLEKSGVFGEGVGNIANFELDPQYLPQOTFKYTIASKDYP	180
DB	488	EKKLGINMLVKLITPDGKVKLEKSGVFGEGVGNIANFELDPQYLPQOTFKYTIASKDYP	547
QY	181	EVSVDGTFVTPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA	240
DB	548	EVSVDGTFVTPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA	607
QY	241	YLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVIVEVPILEKENQTD	300
DB	608	YLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIVIVEVPILEKENQTD	667
QY	301	KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSTLEEVEPTVDPVOE	360
DB	668	KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSTLEEVEPTVDPVOE	727
QY	361	KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFTGEAPQNGENKPSNGKV	420
DB	728	KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFTGEAPQNGENKPSNGKV	787
QY	421	STGTVENQPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGVSDPMLDPALEAPAV	480
DB	788	STGTVENQPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGVSDPMLDPALEAPAV	847
QY	481	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	528
DB	848	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	895
RESULT 14			
AAU84080	ID	AAU84080	standard; Peptide: 896 AA.
XX	AC	AAU84080;	
XX	DT	08-MAY-2002	(first entry)
XX	DE	S. pneumoniae derived chimeric peptide, VP123.	
XX	KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;	
XX	KW	pneumonia; streptococcal bacterial infection; mutant; mutain;	
XX	KW	BVH-11-2.	
XX	OS	Streptococcus pneumoniae.	
XX	OS	Synthetic.	
XX	PN	WO200198334-A2.	
XX	PD	27-DEC-2001.	
XX	PF	19-JUN-2001; 2001WO-CA00908.	
XX	PR	20-JUN-2000; 2000US-212683P.	
XX	PA	(SHIR-) SHIRE BIOCHEM INC.	
XX	PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
XX	DR	WPI; 2002-122272/16.	
XX	PS	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia	

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX

SQ Sequence 896 AA:

Query Match 100.0%; Score 2746; DB 23; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.7e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPIDHKPVGIGHSH 60
DB 369 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPIDHKPVGIGHSH 428

QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKSTNNONFTLANGOKRVSFPPEL 120
DB 429 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKSTNNONFTLANGOKRVSFPPEL 488

QY 121 EKKGINMLVKLITPDGKVLKSVKFGVGNFIANFELDQYLPFGOTFKYTIASKDYP 180
DB 489 EKKGINMLVKLITPDGKVLKSVKFGVGNFIANFELDQYLPFGOTFKYTIASKDYP 548

QY 181 EYSYDGTFTVPSLAYKASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVDFEFHNA 240
DB 549 EYSYDGTFTVPSLAYKASQTIFFPHAGDTYLRVNPQFAVPKGTDLALRVDFEFHNA 608

QY 241 YLENNKYGEIKLIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEPILKENQTD 300
DB 609 YLENNKYGEIKLIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEPILKENQTD 668

QY 301 KPSILPQKRNKAQENSKLDEKVEPKISEKVEKELSETGNSTNSNLEEVPIVDPVQE 360
DB 669 KPSILPQKRNKAQENSKLDEKVEPKISEKVEKELSETGNSTNSNLEEVPIVDPVQE 728

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYPSGVEIKKNMADFTGEAPQNGENKPSGENKV 420
DB 729 KVAKFAESYGMKLENVLFNMDGTIELYPSGVEIKKNMADFTGEAPQNGENKPSGENKV 788

QY 421 STGTVENQTEKNKPADSLPEAPNEXPKVPENSTDNGLNPEGNVSGSDMLDPALEAPAV 480
DB 789 STGTVENQTEKNKPADSLPEAPNEXPKVPENSTDNGLNPEGNVSGSDMLDPALEAPAV 848

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSPFIA 528
DB 849 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSPFIA 896

RESULT 15
AAU84068
ID AU84068 standard; Peptide; 901 AA.
XX
AC AU84069;

XX 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, VP110.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; mutein; BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO2001198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

PR (SHIR-) SHIRE BIOCHEM INC.

PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia.

XX Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX

SQ Sequence 901 AA:

Query Match 100.0%; Score 2746; DB 23; Length 901;
Best Local Similarity 100.0%; Pred. No. 4.7e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPIDHKPVGIGHSH 60
DB 374 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPIDHKPVGIGHSH 433

QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKSTNNONFTLANGOKRVSFPPEL 120
DB 434 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKSTNNONFTLANGOKRVSFPPEL 493

QY 121 EKKGINMLVKLITPDGKVLKSVKFGVGNFIANFELDQYLPFGOTFKYTIASKDYP 180
DB 494 EKKGINMLVKLITPDGKVLKSVKFGVGNFIANFELDQYLPFGOTFKYTIASKDYP 553

```

QY 181 EYSYDGTFTVPTSLAYKMASQTFYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEFGNA 240
Db 554 EYSYDGTFTVPTSLAYKMASQTFYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFDEFGNA 613
QY 241 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTD 300
Db 614 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTD 673
QY 301 KPSILPQFRKRAQENSKLDEKVEEPKTSKVEKESKLSKTSNSTLEEVPTVDPVQE 360
Db 674 KPSILPQFRKRAQENSKLDEKVEEPKTSKVEKESKLSKTSNSTLEEVPTVDPVQE 733
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEVIRKKNMADFTGEAPQNGENKPSSENGKV 420
Db 734 KVAKFAESYGMKLENVLFNMDGTIELYLPSEVIRKKNMADFTGEAPQNGENKPSSENGKV 793
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNGVSDPMLDPALEAPAV 480
Db 794 STGTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNGVSDPMLDPALEAPAV 853
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
Db 854 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 901

```

Search completed: May 13, 2003, 13:53:34
Job time : 23.5622 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:27 ; Search time 11.9447 Seconds

(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2746

Sequence: 1 MKDLDDKTEERKIAIMKQYG.....IELRLPSEVIKKNSDFIA 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2740	99.8	1039	2 D97985	hypothetical prote
2	2734	99.6	1039	2 H95115	conserved hypotet
3	184	6.7	2004	2 F95133	immunoglobulin A1
4	169.5	6.2	1963	2 B98002	IgA-specific metal
5	153	5.6	1873	2 T30944	surface protein pr
6	149.5	5.4	4688	2 F92885	hypothetical prote
7	149	5.4	1849	2 C41859	IgA-specific metal
8	148	5.4	1939	2 T18372	repeat organellar
9	146.5	5.3	791	2 T39924	hypothetical prote
10	146	5.3	1272	2 C90593	hypothetical prote
11	143.5	5.2	1139	1 E64234	cytadherence-acces
12	142	5.2	2485	1 H71621	serine/threonine-s
13	141.5	5.2	1441	2 B66807	hypothetical prote
14	141.5	5.2	1664	2 T18262	S-layer protein -
15	141	5.1	1134	2 A60234	IgA FC receptor pr
16	141	5.1	1164	1 FCSOAG	IgA FC receptor pr
17	139	5.1	750	2 G41361	probable flagellin
18	138.5	5.0	1127	2 T25804	hypothetical prote
19	137.5	5.0	387	2 G37011	alanyl-tRNA synthe
20	137.5	5.0	1526	2 A44406	DNA topoisomerase
21	137	5.0	940	2 A31744	internalin protein
22	136.5	5.0	940	2 A31744	internalin protein
23	134.5	4.9	2481	2 D90011	FmtB protein [impo
24	134	4.9	380	2 S22415	membrane protein T
25	133	4.8	710	2 S30154	low-temperature-in
26	133	4.8	1020	2 T29108	hypothetical prote
27	132.5	4.8	858	2 S15762	neurofilament trip
28	132	4.8	1231	2 T28661	rib protein - Stre
29	131.5	4.8	1701	2 A26868	major merozoite su

30	131	4.8	774	2 A24057	glycophorin-bindin
31	130.5	4.8	719	2 AE1131	hypothetical prote
32	130.5	4.8	853	2 C97985	hypothetical prote
33	130.5	4.8	896	2 AB1156	conserved membrane
34	130.5	4.8	1185	2 A42404	collagen adhesin -
35	130.5	4.8	1750	2 H64403	ribonucleoside-tri
36	130.5	4.8	5005	2 F82884	hypothetical prote
37	130	4.7	1702	2 A41859	IgA-specific metal
38	130	4.7	1983	2 G86643	hypothetical prote
39	129.5	4.7	855	2 D98004	histidine Motif-Co
40	129.5	4.7	1086	2 S16752	major merozoite su
41	129.5	4.7	1094	2 S49313	protein kinase - s
42	128.5	4.7	366	2 S56040	pheromone response
43	128.5	4.7	384	2 G86287	hypothetical prote
44	128.5	4.7	933	2 S41539	fibrinogen-binding
45	128.5	4.7	941	2 F97353	uncharacterized co

ALIGNMENTS

RESULT 1

D97985

hypothetical protein phrE [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: D97985

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

Y, P.; Sun, P.N.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: D97985

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1039 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:gl5458515; GSPDB:GN00174

C:Genetics:

A:Gene: phrE

Query Match 99.8%; Score 2740; DB 2; Length 1039;

Best Local Similarity 99.8%; Pred. No. 1.8e-150;

Matches 527; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKDLDDKTEERKIAIMKQYGVKRESIVVVKENAIIPHGDDHADPTDEHKPVGIGHSH 60

Db 512 MKDLDDKTEERKIAIMKQYGVKRESIVVVKENAIIPHGDDHADPTDEHKPVGIGHSH 571

Qy 61 SNYELFKPEEGVAKKGNKYVTGEELTNVNLKNSTFNNONFTLANGQKRVSFSPPEL 120

Db 572 SNYELFKPEEGVAKKGNKYVTGEELTNVNLKNSTFNNONFTLANGQKRVSFSPPEL 631

Qy 121 EKKLGINMLVLIIPDGKVLKGVKGVGAGVNIANFELDPQVLPQGTFFYTTASKDYP 180

Db 632 EKKLGINMLVLIIPDGKVLKGVKGVGAGVNIANFELDPQVLPQGTFFYTTASKDYP 691

Qy 181 EVSYDGTFTVPTSLAYKNASQTIFFYPFHAGDTYLRVNPQFAVPKGTALVRVDFEHGNA 240

Db 692 EVSYDGTFTVPTSLAYKNASQTIFFYPFHAGDTYLRVNPQFAVPKGTALVRVDFEHGNA 751

Qy 241 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVEPILKEKQTD 300

Db 752 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVEPILKEKQTD 811

Qy 301 KPSIIPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSQTSNSTSNSTLEEVPTVDVQVE 360

Db 812 KPSIIPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSQTSNSTSNSTLEEVPTVDVQVE 871

Qy 361 KVAFAESYGMKLENVLNFMNDGTIELYLPSEVIKKKNADFTGEAPOGNGENKPSYENKGV 420

Db 872 KVAFAESYGMKLENVLNFMNDGTIELYLPSEVIKKKNADFTGEAPOGNGENKPSYENKGV 931

```
RESULT 3
F95133
immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95133
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75263.1; PID:g14972632; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
C:Gene: SPl154
C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 6.7%; Score 184; DB 2; Length 2004;
Best Local Similarity 23.5%; Pred. No. 0.023;
Matches 130; Conservative 70; Mismatches 162; Indels 192; Gaps 34;

QY 60 HSNYELFKPEGVAKKGNKYT-----GEELTNVYN-----LLKNSTFN-----NQN 102
Db 279 HKNLET-KKEKISPKETGVNTLNPQDEVLSGOLNKPPELLYREETMETKIDFEEIQEN 337
QY 103 FTLANGKRVSFSPPELEKKGILNM-LVKLIT-----PDGKVLKVS 144
Db 338 PDLAEGTVRV-----KOEKLGKKEIVRIFSVNKEEVSEIVSISTTAPSPRIVEKGT 391
QY 145 GK--VFGE--GV-----GNIANFELDQYLPQGTFFKTYTASKDYPEVSDGTFITVP 191
Db 392 KKTQVIREQPTGVGHKDVQSGAIVEPAI-QPELP-----EAVVSDKGEPEVQP---TLP 442
QY 192 TSLAYKMASQTIFFPHAGDTYLYRVNQ-----FAVPKGTALRVFDEFHGNAYLENNYK 247
Db 443 EAVV-----TDKGET--EVQPESTVTVSDGPEQPEVAPLPEYKGN-----481
QY 248 VGEIK-LPIPKL-NOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKQTDKPSI 304
Db 482 IEQVKPEPTVEKTEQGEPEKT--EEVPV-----KPTETPVNPNNEGTEGTSI 527
QY 305 LPQKRN--KAQNSKLDEKVEEPKTEK-----VEKEKLESETGS 343
Db 528 --QEAENPVQAEESTTNSEKV-SPDTSSKNKTGEVSSNPSTSTSVGESNKPENHDSKNE 584
QY 344 TSNSTLEEVPTVPVQEKVAKFAESYGMKLENLFNMDGTIELYLPSEVGIKKMADFTG 403
Db 585 NSEKIVEVP-VNP-----NEGIVE-----GTSNQET 610
QY 404 EAPQNGENKPSSENGKV---STGTVENQPTENKP--ADSLPEAPNEKPVKPEPS---TD 454
Db 611 EKPVPQAEETQNTSGKIANENTGEVSNKPSDSKPPVEESNQPEKNGTAIKPENGSTTSE 670
QY 455 NGMLNPE--GNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYGLGLSDVFNNMDGTIEL 511
Db 671 NGQTEPEFNSNGNSTEDVSTESNTSNGNEIKQENE-----LDPDKKVEPEKTELEL 723
QY 512 RLPSGEVKKNLSD 525
Db 724 R-----NVSD 728

RESULT 4
B98002
IgA-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
```

```
RESULT 2
H95115
conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
C:Gene: SPl1004

Query Match 99.6%; Score 2734; DB 2; Length 1039;
Best Local Similarity 99.6%; Pred. No. 3.9e-150;
Matches 526; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVRESIVVWKNKNAIYPHGDDHHADPDIDEHRPVGIGHSH 60
Db 512 MKDLKKIEKTAGIMKQYGVRESIVVWKNKNAIYPHGDDHHADPDIDEHRPVGIGHSH 571
QY 61 SNYELFKPEGVAKKGNKYTGEELTNVYNLLKNSTFNQNTFLANGKRVSFSPPEL 120
Db 572 SNYELFKPEGVAKKGNKYTGEELTNVYNLLKNSTFNQNTFLANGKRVSFSPPEL 631
QY 121 EXKLGINMLVKLITPDGKVLKGVGEGVGNITANFELDQYLPQGTFFKTYTASKDY 180
Db 632 EXKLGINMLVKLITPDGKVLKGVGEGVGNITANFELDQYLPQGTFFKTYTASKDY 691
QY 181 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLYRVNQFOFVAPKGTDALRVFDEFHGN 240
Db 692 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLYRVNQFOFVAPKGTDALRVFDEFHGN 751
QY 241 YLENNYKVGEEKILPIPKLNGQTTTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEK 300
Db 752 YLENNYKVGEEKILPIPKLNGQTTTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEK 811
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPKTEKVEKELSETGNSSTSTLEEYPTVPDQVE 360
Db 812 KPSILPOFKRNKAQENSKLDEKVEEPKTEKVEKELSETGNSSTSTLEEYPTVPDQVE 871
QY 361 KVAKFAESYGMKLENLFNMDGTIELYLPSEVGIKKMADFTGEAPQNGENKPSSENGKV 420
Db 872 KVAKFAESYGMKLENLFNMDGTIELYLPSEVGIKKMADFTGEAPQNGENKPSSENGKV 931
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPEPSNSTDNGMLNPEGVGSDPMLDPALEAPAV 480
Db 932 STGTVENQPTENKPADSLPEAPNEKPVKPEPSNSTDNGMLNPEGVGSDPMLDPALEAPAV 991
QY 481 DPVQEKLEKFTASYGLGLSDVFNNMDGTIELRUPSGEVKKNLSDIFA 528
Db 992 DPVQEKLEKFTASYGLGLSDVFNNMDGTIELRUPSGEVKKNLSDIFA 1039
```

C:Accession: B98002
E: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A: Title: Genome of the Bacterium Streptococcus pneumoniae strain R6.
A: Reference number: A97872; MUID: 21429245; PMID: 11544234
A: Accession: B98002
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1963 <KUR>
A: Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:gl5458662; GSPDB:GN00174
C: Genetatics:
C: Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C: Keywords: hydrolase, metalloproteinase

Query Match 6.2%; Score 169.5; DB 2; Length 1963;
Best Local Similarity 23.7%; Pred. No. 0.16;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY 60 HSNYELFRPEGVAKKGNKYVT---GEELTNVYN---LLKNSTFN---NON 102
DB 279 HKNLET-KKEKISPKKXTGNTLNPODEVLSGQLNKPPELLYRETIETKIDFQEEIOEN 337
QY 103 FTLANGQKRVSESPPELEKLGIM-LVKLIT-----PDGKVLKVS 144
DB 338 PDLAEGTVRV-----KQEGKLGKVEIVRIFSVNKEEVSREIVSTSTTAPSPRIVEKGT 391
QY 145 GK--VEGE---GV-----GNTANFELDOPYLPQGTFKYTIASKQYPEVSYDGTETVP 191
DB 392 KKTQVKEQPETGVHEKDVQSGAIVEPAI-OPELP-----EAVSCKGEPEVQ---TLP 442
QY 192 TSLAYKMASQTIFFPHAGDTYLRVNPQ----FAVPKGTDALRVDFEFGHAYLENNYK 247
DB 443 EAVV-----TDKGET--EVOPESPDTVSDKGEQVAPLPEYKGN-----481
QY 248 VGEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSYIIVFPILLEKENOTDKPSI 304
DB 482 IEQVAPETVPVTEKQGEKPT--EVPV-----KPTETPVNPNEGTETGSI 527
QY 305 LPQKRN---KAQENSKLDERVEEPKTSKV-----EKEKLSSETGNT 344
DB 528 --QEAPENVPAPAEETTNSEKVSPTDTSSENTGEVSSNPDSSTTSVGSNKPEHNSKNEN 585
QY 345 SNSILEEYPTDVPQEVAKFAESYGMKLENVFNMDGTIELYLPSEGVKKKNMADFTGE 404
DB 586 SEKTVEEVP-VNP-----NEGTVS-----GTSNQETE 611
QY 405 APQNGENKPSNGKV---STGTVENQPTENKP--ADSLPEAPNEKPKVPKPSN---TDN 455
DB 612 KPQVPAETQTSNGKIANENTGEVSNKPSDSKPPVEESNQPEKNGTKATKPPNSCNTTSEN 671
QY 456 GMLNPE 461
DB 672 GQTEPE 677

RESULT 5
T30944
surface protein precursor - Enterococcus faecalis
C: Species: Enterococcus faecalis
C: Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C: Accession: T30944
R: Shankar, V.; Baghdavan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A: Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene end
A: Reference number: 220943; MUID: 99081742; PMID: 9864215
A: Accession: T30944
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1873 <SHA>

A: Cross-references: EMBL:AF034779; NID: g3873186; PID: g3873187; PIDN: AAD09858.1

Query Match 5.6%; Score 153; DB 2; Length 1873;
Best Local Similarity 23.7%; Pred. No. 1.3;
Matches 123; Conservative 58; Mismatches 194; Indels 144; Gaps 28;

QY 2 KLDKKTI---EETIAGIMKQYGVKRESIVVNEKNAIYPHGDHHADPIDEHKPVGIG 57
DB 1027 KDLPKDAKYTWKEVD--ISTAGKKGTVVVT-----YSDGSSDEVE-VD---VTVT 1072
QY 58 HSHSNYELFKPE-EGVAKKGNKYTYGEELTNVNNLKNSTFNN-----NQFTLANGQ 109
DB 1073 DNRSDADKYEPTVEGEKVEIGKVDLTNDVNTLPTLPOGTTVDVTPGGTIDTNTPGNVE 1132
QY 110 KRVSESPPELEKKGIMLVKL---ITPDGKVLKVSQKVFGEV-----GNIAN 157
DB 1133 GVIEVTPD-----GTRKDTVKVPVEVTDNRSDADKYEPTVEGEKVEIGKVDLTNTVTN 1186
QY 158 FELDQPYLPQGTFKYTIASKQYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVN 217
DB 1187 L-----PFLPGT---TVT-----DVTGGTIDTNTPGNEGVIEV-----TV--- 1221
QY 218 PQRAVPKGTDALRVDFEFGHAYLENNY-----KV---GEIKLP-----TPKLNQGT 262
DB 1222 -----PDGTRDTVKVPVEVTDNRSDADKYEPTVEGEKVEIGKVDLTNDVNTLPTLPGT 1276
QY 263 T---RTAGNKIPVTFMANAYLDNQSY-----IVEVPILLEKENOTDKPSILPQFKRNKA 313
DB 1277 TVTDVTPGGTIDTNTPGNEGVIEVTPDGTQKDTVKVPVEVTDNRSDADKYTPMVEGEKV 1336
QY 314 QENSKLDERVEEPTSEKVEKEKLSSETGNTSISTLEEYPTDVPQEVAKFAESYGMK 373
DB 1337 EIOGKVD-----LTDNVNTLPTLPOGTTVDVTPG-----GTID 1370
QY 374 ENVLNFMNMGDTIELYLPSE--EVIK-----KNMADTGEAPQNGENKPSNGKVS-TGT 424
DB 1371 TNPNGNEGVIEVTPDGTQKDTVKVPVEVTDNRSDADKYEPTVEGE-KVEIGKVDLTND 1429
QY 425 VENQPTENKPADSLPEAPNEKPKVPKPSNDTNGMLNPEGN 463
DB 1430 VTNLPT-----LPGTTVTDVTPGGTIDT---NTPGN 1458

RESULT 6
F82885
hypothetical protein U0482 [imported] - Ureaplasma urealyticum
C: Species: Ureaplasma urealyticum
C: Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C: Accession: F82885
R: Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A: Description: the complete sequence of Ureaplasma urealyticum: Alternate views of a
A: Reference number: A82870
A: Accession: F82885
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-4688 <GLA>
A: Cross-references: GB:AE002145; GB:AF222894; NID: g68999476; PIDN: AAF30894.1; GSPDB: GN
A: Experimental source: serovar 3; biovar 1
C: Genetatics:
A: Gene: U0482
A: Genetic code: SGC3

Query Match 5.4%; Score 149.5; DB 2; Length 4688;
Best Local Similarity 20.6%; Pred. No. 8.1;
Matches 125; Conservative 89; Mismatches 219; Indels 173; Gaps 27;

QY 1 MKLDKKIEKIKIAGIMKQYGVKRESIVVNEKNAIYPHGDHHADPIDEHKPVGIGHSH 60
DB 3859 INDTQQTINVTLSGVNSKYN-GRQIKVYKDNVNIY-----ESSLITLQKQK 3905
QY 61 SNVELFKPEEGVAKKGNKYVTGTEELTNVNNLKNSTFNNQFTLANGQKRVSESPPEL 120

3906 NDYQLL-----LSNLNSNRYEYREKIE--INHISN-TNNEFDEKLANG---VSNTFITQT 3954
QY 121 EKKL-----GINMLVKLTPDGKVKLEKSGKVGFGVGNANFELDQYLP 166
DB 3955 KNTTQWNDSSATIIVGTGVAFNFKIKSED-KILEN-----NQQVAVFAP 3999
QY 167 GQTFK-----YTLASRDYPEVSYDGFVPTSLAYKMASOTIFYPFHAGDTYLRNPQ 219
DB 4000 RETIRDTWLTQYTRPLKDYVSDFKEGTWAHDLNSVNFKEET-----TYKLVKIQ 4050
QY 220 FAVPFGTDLVRVDFEFGHAYLEN-----NY-----KVGEIKL-PIPKLNOQTTRT--- 265
DB 4051 F-VNKPTRAKNNI-NNSENNVLDTNSINSYEFTRKVGDKHLNITSSNNVNTSOTI 4108
QY 266 -----AGNKIPVTFMANAVLDNQSTYIVVEPILKEKNOTDKPSILPOPKRKAQ 314
DB 4109 NTLGSKVKSWSGKKIKLUSKN-----DTSEIHTNEVLIESNKTQYNILLNNKRNRTY 4164
QY 315 ENSKLDERVEEPEKTSERKEKLSGTSTGNSSTLEEVPTVDPVOEKVAKFAESYGMKL- 373
DB 4165 --TLIDVKLIDNNVSDFPKGNLTNSFIITRTSAINVLNIEEISNRASTNLKSTIIKIN 4222
QY 374 ----ENVLFMDGTIELYLPSEVITKKMAVDTGEAPOGNGENKPSNGKYSTGT----- 424
DB 4223 LNDPDLVLRDKDOATIV-----GNKKQAMGFITVSGNIKYLTAIVLDLN 4267
QY 425 -----VENQPTENKPA-----DSLPEAPNEKPKPENSTD---NGM 457
DB 4268 FNDKYNIVNISFNKPSIAAENIGDKSNIIYNNDSP-----KLEINNDIIVNGP 4319
QY 458 LNPE-----GNVGDPMLDPALEAPADVPQVEKLEKTASYGLD-----SVIFNMD 506
DB 4320 INKEIYVKNANQKNIDVDLGLQINPKIAHLNRTFAKEFKSTNNDIETNVINGSLVND 4379
QY 507 GTIEIR 512
DB 4380 GTSIR 4385
RESULT 7
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
C:Species: Haemophilus influenzae
A:Variety: strain HK613
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: C41859
R:Poulsen, K.; Reinholdt, J.P. Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A:Reference number: A41859; MUID:92234949; PMID:1373717
A:Accession: C41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1849 <PO>
A:Experimental source: strain HK613
A:Note: sequence extracted from NCBI backbone (NCBIP:97285)
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
Query Match 5.4%; Score 149; DB 2; Length 1849;
Best Local Similarity 20.1%; Pred. No. 2.2;
Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
QY 28 VAKENKAIIPYHCDHHADPIDHKPVGIG-----HSHSNYELFKPEGVAKK----- 75
DB 705 LNVKGTFLSGRPTPHARDI-----AGISSTKDPHTENNEVVVDDWINEFKATIM 759
QY 76 --EGN-KVYTGEBLTNVNLLKNSFNNOFTLANGQKRVSFSPPELEKKLGINMLVK- 131
DB 760 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIQY-----KTGDTVCVRS 802
QY 132 ----LITPDGKLVLEKSVKVGEGV-GNI-----ANFELDQYLPFGOTFKYTIASKDYP 180

803 DYTGYVTCHNSNLSEXALNSFNPTNLRGVNVNTENASFTLGRANLFG-----TIGSIGTS 857
QY 181 EVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNA 240
DB 858 QVN-----LKENSHWHLTGNSNVNOLNLTNGHILHNAONDANKVTTYNTLVNSLSNG 911
QY 241 -----YLENNYKVEIKLPI-----PKLNQGT-----TRTAGNKIPVTFM 275
DB 912 SFYVYVDFTNKSNKVVAKSATGNFTLQVADKTGEPNENELTFEDASNATRNNEVT-L 970
QY 276 ANAYLDN-----QSTYIVVEPILKEKNOT-----DKPSILPOPKRN 311
DB 971 ANGSVORGAWKYKLRNVNGRYDLYNPEVEKRNQTVDTTNTITPNDIQADAPS----- 1022
QY 312 KAQENSKLDEKVEEP-----KTSKEVEKEKJSE 339
DB 1023 -AOSNNEIARVETPVPPAPAPATESAIASEQPETRPAETAQAPEMETNTANSTETAPKSD 1081
QY 340 TGNSTNSTLEEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVKKNMA 399
DB 1082 TATOTENPNSVESVPS-----ETTEKVAENPPQENETVAKNEQENTPTFQNGEVAKEQD 1136
QY 400 DF-----TGEAPOGNGENKPSNGKYSTGTVENQPT-----ENKPADSLPEAPNEKPY- 447
DB 1137 TVANTQTNEATOSEGKTEETQ-----TAEKSEPTESVTSNQPEKTVSQSTEDKVVV 1191
QY 448 ----KPNSTDNGLNAPGNGVSDP-MLDPALEEAFAVDPVQE 485
DB 1192 EKBEKAKVETEETQKAPQVTSKEPPKQAEPAPEEVTDTNAEE 1234
RESULT 8
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: EMBL:U043145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1
Query Match 5.4%; Score 148; DB 2; Length 1939;
Best Local Similarity 19.3%; Pred. No. 2.7;
Matches 117; Conservative 101; Mismatches 206; Indels 182; Gaps 27;
QY 4 LDKKIEEKIAGIMKQYGVKRESIIVNKNKAIYPHGCHHH---ADPIDHKPV----- 54
DB 989 LGEEHKVWAGIEEKY--KVEATKLAEEHKDVTYKLGQEKKEIAKLEDGKHEVNEVEK 1046
QY 55 -----GIGHSHSN-----YELFPEEGVAKKGNKVYTGEEELTNV 90
DB 1047 KNASLLNMLENHNKEMIKLKEEHKESASDLVEKLYQKDEEV-KNSNNKI---EELTNVI 1102
QY 91 NLLKNSFTNNQNFTRANQKRVSFSPPELEK-KLGNMLVKLITPDGKLVKSVG-KVF 148
DB 1103 KDLNDSIMCKYKQILEEVEKRNENY--FEINKLIVQNEKMDM--NDKKLLEKENEIKKL 1158
QY 149 GEGVGNIANFELDQYLPFGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIYPFH 208
DB 1159 NKLSNKKVPETKE-----NTYK----- 1176
QY 209 AGDTYLRVNPQFAVPKGTDALV--RVDEFHGNAYLENNYKVGEEK--LPIPKLNOQTTR 264
DB 1177 -----NSMWYNNENKERTIIVDSCKENISESDVEG--KGNLKMFTLSKKERNIFS 1226
QY 265 TAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNOTDKPSI-----LPQF 308

Db 1227 INDNKESSELVDTI---KSAVINKIEMVKKIEDNGKNIEDLKKKILDLNSELINLENA 1283
QY 309 KRKAQENSKL-----DEKVEEPTKEVP-----KELSETGNSTNSYLE 350
Db 1284 KWLVDENNKKETEIKDNKLNKEKKNENTEILNLDIIKLKKEISEWKEDEEKLKE 1343
QY 351 EYPTVDVPOEVKAKFAESYGMKLENVL--FN-----MDGTIELYLPSEVIVKKNMAD 400
Db 1344 NI-----KLANDIEQINKVEKKEEMLKFNENINEVTSLNQIEKMKLELNKVEL 1399
QY 401 FYGEAPOGENKPKSPENKSVTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNP 460
Db 1400 LIAEKRETNMSISNDKNKIVENILED--TDSK-----QNNLNK---NV 1438
QY 461 EGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGL-----DSVIFNMDG---TIEL 511
Db 1439 EDKTGDDINCKENNDQAREISVLKDEIKISMLYGEELNKRNSYDEKVKNTNELKEIKI 1498
QY 512 RLPSGE 517
Db 1499 RNKKG 1504

RESULT 9
T39924
hypothetical protein SPAC21B10.03c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39924
R:McDougall, R.C.; Rajadream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, October 1997
A:Reference number: 221891
A:Accession: T39924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-791 <MCD>
A:Cross-references: EMBL:AL121794; PIDN:CAB57927.1; GSPDB:GN00067; SPDB:SPAC21B10.03c
A:Experimental source: strain 972h; cosmid c21b10
C:Genetics:
A:Gene: SPAC21B10.03c; SPDB:SPAC21B10.03c
A:Map position: 2

Query Match 5.3%; Score 146.5; DB 2; Length 791;
Best Local Similarity 22.1%; Pred. No. 0.86;
Matches 115; Conservative 67; Mismatches 193; Indels 155; Gaps 26;

QY 46 DPIDEKPV-GIGHSHSNVELFKPEGVAKKEGNKVYTGEEELTNVNLKLNSTNNONFT 104
Db 126 DPEDAGVPLAGL-----EESTDNVEWQDFATNEKLFV-----KSHFEDELYT 168
QY 105 LANGOKRVSFSPPPPELEKGLGINMLVK---LITPDGKLVLEKSVKVFGEV-----G 153
Db 169 -----SRIDRSHPKYKEKEQADRIAKIEGTVINNHIAERGLKVDSDGLDEEDLYSG 223
QY 154 NTANFELDQPLPGQTFKYITASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTY 213
Db 224 VERSDVVRNT--RSNAYNNKNDQKPKNHAPHQHP-----QKVVVPP---DDPA 270
QY 214 LRVNQFAVPK--GTD--ALRVFDEPHGNAYLENNYKVGSIK-----LPTKLNQ 261
Db 271 IVSHRLALPRAGPDSRAAEFFNARRKAGPLSRREKEGQIKFQMPSQSLKGTGSLD-- 328
QY 262 TTRTAGNPIPTFMANAYLDNQSTVIVEVPILEKENQNDKPSILPQFKRNKAQENSKLDE 321
Db 329 -----SKQP-----SSTKSAEVKVADEKQLPDASS-----QATADSKEPKE 367
QY 322 KYEPEKT-----SKVKEKELSETGNSTNSLTLEEVTVPDVQEK-VAKFAESYGMKL- 373
Db 368 EAEKPVTSATEVSSEKVEK---VDGNTSSPSKEEKSPSTEPEKPSVVYTORKETGTGKL 423
QY 374 -----ENVLFNMDGTIELYLPSEVIVKKNMADFTGEAPQ----- 407
Db 424 TKLNKATISFRPNVAPVFTGKFTIPSKPAPVNASRPMPPQSQNNSEASIPSTTPSPS 483

QY 408 --GNGENKPS-----NGKYST-----GTVENOPTENKPAD-----SLPEAPNEK 445
Db 484 VVNGENKPSPPVFFRQPVSSSEKEPILDNFNVFNKNGEEHQAEQIDKPFSCPTWTNTG 543
QY 446 PVKPNSTNGMLN--PEGNVGSDPMLDPALEAPAVDPV 483
Db 544 P-----NSLQQTIANRPEGNGSGS-----AKKAAANPM 572

RESULT 10
C90593
hypothetical protein MYPV.6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: C90593
R:Chamcaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: C90593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1272 <KUR>
A:Cross-references: GB:AL445566; PID:G14090066; PIDN:CAC13824.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV.6510
A:Genetic code: SGC3

Query Match 5.3%; Score 146; DB 2; Length 1272;
Best Local Similarity 20.3%; Pred. No. 1.9;
Matches 125; Conservative 94; Mismatches 238; Indels 158; Gaps 28;

QY 1 MKDLDDKIEEKIAGIMQYGVKRESIVVNEKNAILIYPHGDHHDHAFIDEHKPVGTGCHSH 60
Db 86 LTDTDDTDEKINVSX-----LISNEQK-----QEENSQRNSQN 121
QY 61 SNYELFKPEGVAKKEGNKVYTGEEELTNV-----NLLKNSTFNQNTLIANGQRVSF 114
Db 122 QNEKFANDDEGTALKVLEFVSYNGVPTREVKLHGFKNTIQKTSNNLNLNLSNKEQ--- 177
QY 115 SFPPEL---EKKLGINMLVK--LITPDGKLVLEKSVKVFGEVGNIANFELQPYLPQG 168
Db 178 IHPSELDSNQQLVTTNFEKLSLETLIDIEKYVLYEVYVNDIDAEAKIRLTLQKLTNE 237
QY 169 TFK--YTIASKDXPEVSYDGT-----TVPTSLAYKMASQTIFFPHAGDTYLRVNP-- 218
Db 238 NIKNSFEVLKDPKKYKLDSDSLKAKINLVKNDLSIEELLNITSKGFANKTEDOYNKLG 297
QY 219 ---QFAYPKGTDALRVFDEPHGNAYLENNYKVGEL--KLPIKLNQGTTRTAGNIPVT 273
Db 298 ETLNIDLPNGYEF-----EFVSLAPKANDASVGLLTYKLVKNLNDGTSNENPSK---- 347
QY 274 FMANAYLDNQSTVIVEV--ILEKENQTDKPSILPQFKRNKAQENSKLDEKVEPEKTSK 331
Db 348 ---NGRIET--NTIVEETINLSLKNDSKDEP-----KDNSNTNSKDEPKTDEPKVEEP 397
QY 332 VEKE-KLSETGNSTNSLTLEE-----VPTVDVPOEKVA----- 363
Db 398 REDEPKTNPSDSKDKDKEIDPENEKPKDQPKTEPKNEPKDKPKVEPKDETLAIFDKISKI 457
QY 364 -----KFAESYGMKLEN--VLFNMDGT--IELYLPSEVIVKKNMADTGEA 405
Db 458 ELAKNSOLKQKLSQFKRES--DLNLSNLKVLVSODKNNKFSLELPQGYSTSFKLA----- 510
QY 406 PQNGENKPSNGKSVSTGTV-----ENQPTENKPADSLPEAPNEKPVKPNSTNGMLN 459
Db 511 -----SNSNDEGTLVDKVIQVQKGEVKRKELKLTNLTFFESLK-----ESDQLDFSN 561
QY 460 PEGNVGSDPMLDPALEAPAVDPVQEK-----LEKFTASYGL-GLDSVIFNMDGTIELR 512
Db 562 KKDPLASSVVNDKIITRESLV--VKNKTIENFDNFKYDISYSVSSSLDEV----NGLKIK 615

B86807
hypothetical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86807
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s.s.
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86807
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:g12724450; PIDN:AAK05556.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yoiC

Query Match 5.28; Score 141.5; DB 2; Length 1441;
Best Local Similarity 22.98; Pred. No. 4.1;
Matches 113; Conservative 52; Mismatches 185; Indels 143; Gaps 29;

Qy 82 TGEELTNVNLKSTNNQFTLANGQKRVSFSPPELEKKGILNMLVLIPT-DGK-- 138
Db 955 TGASISSVANITVNDSSNTTPTDGTGGNTT-----NPTDGNIT 994

Qy 139 VLEKSGKVFEGVGNANTANFELDPYLPQTFKTYIASK-----DYPEVSYD---GTFTVP 191
Db 995 VLPQDQGLK---GTADDVTVTDQPLSPGSDGVTLPSSDGKVRDPGGSYVPGTVDVP 1051

Qy 192 TSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDALVRVEDEFGHNAVLENNYKV--- 248
Db 1052 DG-----TIHLP---GGVINPGGSVTP-GPDGKTGTDGDD-----TTLNPNSPVPG 1095

Qy 249 --GEIKPIPKLNOGTRTA-GNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTDKPSIL 305
Db 1096 DNGSVTLT---GGGTASTPNGN---ITLPGTGYVDPDGT---IHLPGGDIVNPDGTTILP 1146

Qy 306 PQFRNKAQENSKLDEKVEEPEKTSKVEKEKLSSTNSTLEEVP---TVDPVQEKV 362
Db 1147 QDQKGTGDDGKI--RPNGFIIDGNSVTLPGGGVITPCTGINTVPGGSVDP----- 1199

Qy 363 AKFAESYGMKLENVFNMDGTIELYLSGSEVVKANMAFTGEAPQGNKPSNGKRVST 422
Db 1200 -----DGTV--HLPGGDIVN---PDGTTLPQDQKGTGTDGKVK- 1235

Qy 423 GTVENOPTENKPADS--LP-----EAPNEKPVKPNST---DNGMLNPEGNV- 464
Db 1236 ---PNGFSINPDGSIILPGGGVITPCTGINTVPGGSVDPDGTIVHLPGGVWNPDGTTIT 1292

Qy 465 -----GSDMLDPALDEAPAVD-----PVQEKLEKFTASYGLDLSVIFNMD 506
Db 1293 LPQDQKGTGDDVNIKP---NGFSVSNPDGSIILPGGGVITNPGGTYNVPGGTIVV-DPD 1348

Qy 507 GTIELPLPSGSEVI- 519
Db 1349 GTI--HLPNGSEVI 1359

RESULT 14
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1664 <FUJ>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 5.28; Score 141.5; DB 2; Length 1664;
Best Local Similarity 21.7%; Pred. No. 5;
Matches 86; Conservative 54; Mismatches 156; Indels 101; Gaps 17;

Qy 117 PPELEKKGILNMLVLIPTDQKLVKSVKVFEGEGVGNIANFELDPYLPQTFKTYIAS 176
Db 597 PSELPDSYVIMELDKTKVREGDV---IIATIRVNNIKNLGAIQIGIKYDPKYLEANFET 553

Qy 177 KD-----YPEVSYDGTFT-----VETSLAYKMASQITFYYPFHAGDTYLRVNPQFAVPKG 225
Db 654 GDPIDEGTTPAVG--GTILKNRDYLDPTGVAINNVSKGIL-----NFA----- 693

Qy 226 TDALVRVEDEFGHNAVLENNYKVGEEKLPIPKLNOGTTT-----TAGNKIPVTEMANAYL 280
Db 694 --AYYVFDYDREEGKSEDTGIIGNIGFVRLKAEDTIRPEELESMPGSDGTMYLMDWL 751

Qy 281 DNOSTIV--EVPILKEKNQ---TDKPS--ILPQFRNKAQENSKLDEKV--EETKTSK 331
Db 752 NRISGVVVIQAPAIKAASDEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 811

Qy 332 VEKEKLSQENSTNSNTLEEVPVDPVQEKVAKFAESYGMKLENVFNMDGTIELYLP 391
Db 812 PEPITDTPSDPTPSDEPTPSDEPT-----PSD 841

Qy 392 EVIKKNMAFTGEAPQGNKPSNGK--VSTGTVENOPT-ENKPADSLPEAPNEKPYK 448
Db 842 E-----PESDEPTSETPEEPIPTDTSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 888

Qy 449 PENSTNGMLNPGNSGSDPMLD---PALDEAPAVDP 482
Db 889 SDEPTPSE--TPEEPIPTDTSDEPTPSDEPTPSDEPTPSDEPT 923

RESULT 15
A60234
Iga Fc receptor precursor - Streptococcus agalactiae (strain SB35)
N:Alternate names: Iga-binding protein; protein Bac
N:Contains: beta antigen
C:Species: Streptococcus agalactiae
C:Date: 08-Dec-1992 #sequence_revision 08-Dec-1992 #text_change 26-Aug-1999
C:Accession: A60234; S14595; A60230
R:Heden, L.O.; Frithz, E.; Lindahl, G.
Eur. J. Immunol. 21, 1481-1490, 1991
A:Title: Molecular characterization of an Iga receptor from group B streptococci: seq
ents with Iga-binding capacity.
A:Reference number: A60234; MUID:91257158; PMID:2044657
A:Accession: A60234
A:Molecule type: DNA
A:Residues: 1-1134 <HE2>
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521
A>Note: the source is designated as group B streptococcus strain SB35
R:Heden, L.; Frithz, E.; Lindahl, G.
Submitted to the EMBL Data Library, March 1991
A:Description: Molecular characterization of an Iga receptor from group B streptococ
fragments.
A:Reference number: S14595
A:Accession: S14595
A:Molecule type: DNA
A:Residues: 1-1134 <HE2>
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521
A>Note: the source is designated as Streptococcus agalactiae
R:Lindahl, G.; Akerstroem, B.; Vaerman, J.P.; Stenberg, L.
Eur. J. Immunol. 20, 2241-2247, 1990
A:Title: Characterization of an Iga receptor from group B streptococci: specificity f
A:Reference number: A60230; MUID:91055597; PMID:2242758
A:Accession: A60230
A:Molecule type: protein
A:Residues: 'X', 39-48, 'X', 50-52, 'X', 54-56 <LIN>
C:Superfamily: Iga Fc receptor
C:Keywords: Cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein

F:1-37/Domain: signal sequence #status predicted <SIG>
F:138-1134/Product: IGA Fc receptor #status experimental <MAT>
F:199-438/Domain: IGA binding #status predicted <IGAL>
F:439-826/Domain: IGA binding #status predicted <IGAL>
F:827-915/Region: proline-rich repeats
F:916-1101/Domain: cell wall-spanning #status predicted <CWS>
F:1102-1129/Domain: transmembrane #status predicted <TMM>

Query Match 5.1%; Score 141; DB 2; Length 1134;
Best Local Similarity 20.0%; Pred. No. 3.1;
Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

Qy 2 KDLKKIEKI-----AGIMK-QYGVKRESIVNKKENAIYPHGDHHDPI--- 48
Db 240 EDAEVKVEELGKLFSSTKAGLDQEIQEHVAKET---SSEENT---QKVDEHYANSLQNL 293
Qy 49 -----DEHKPVGIGHSHSNVELFKPEEGVAKKEGNKY--TGEELTNVVNL 93
Db 294 AOKSLEELDQATTNEQATQVKNQFNLENAQKLKEIQLIKETNVKLYRAMSESLQVEKEL 353
Qy 94 KNSTFNQNFILANGOKRVSFSPPELEKKGILNMLVKLIITPDGKVLKSVKVGEGVG 153
Db 354 KNSANLEDLVAKSEIVR-----EYEGKL--NOSKNL--PELKQLEEEAHSLKQVVE 404
Qy 154 NT-ANFELDQPYLPQOTFKYTIASKD-----YPE--VSYDG---TFTVPTSLEYK 197
Db 405 DFRKKFKTSEQVTPKKRYKRDLAANNQKQIETLTVSPENITVYEGEDVKFTVTA---K 460
Qy 198 MASQIFIFPFHAGDYLRVNPQFVAPKGTDALRVDFEFGHGNAYLENNKYVGEIKLPIPK 257
Db 461 SDSKT---TLDFSDLLATKYNPSV-----DRISTNYKNT--DNHKIAETIKNLK 506
Qy 258 LNOGTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILEKENQTDKPSILPQPKRN 311
Db 507 LNESQTVILKAKDDSGNVVETK-----TITVQKREEK-----QVEKT 544
Qy 312 KAOENSKLDEKV-EEPKYSERKEVEKLESETGN-----STNSLTLEEVPTV----- 355
Db 545 PEQKDSKTEEKYPQEPKSNKNDKNQLOELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQ 604
Qy 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSGEVIK 395
Db 605 KSINWSQKEPIGEAITSFKKIIGDSSSKYITEHYFNKYKSDFMNYQLHAQM-----EMLT 659
Qy 396 KKNADFTGEAPQNGENKPSNGKVST-----GTVENOPT-----NK----- 433
Db 660 RKVVQYMNKYPDPAETIKKIFESDMKRTKEDNYGSLNDALKGYFEKYEFTLPFNKIKQIVD 719
Qy 434 PADSLPEAPNEKVPKPNSTONGMLNPEGNVSGDPMLDPALEEA-----PAVDPVQ 484
Db 720 DLDKVVEQDQAPI-PENSE-----MQAKEKAKIAVSKYNSKVLGDGVH 762
Qy 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 763 QHLQKKNHSHKIVDLFKLELAIKQQTIFDID 792

Search completed: May 13, 2003, 13:57:16
Job time : 25.9447 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 ; Search time 6.19355 seconds
(without alignments)
3535.857 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2746

Sequence: 1 MKDLDKIEKIKIMQYQ.....IELPLSGEVIKKNLSDFIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	5.4	1849	IGA4_HAEIN	P45386 haemophilus
2	143.5	5.2	1139	HMW1_MYCGE	Q49413 mycoplasma
3	141.5	5.2	1664	SLP1_CLOTPM	Q06852 clostridium
4	141	5.1	1164	BAG1_STRAG	P27951 streptococc
5	140	5.1	1013	SCA4_RICRH	Q9a181 rickettsia
6	137.5	5.0	1526	TP2A_CRIGR	P1515 cricetus
7	134	4.9	357	TPN3_RAT	P19814 rattus norv
8	133	4.8	710	LP78_LARATH	Q06738 arabidopsis
9	132.5	4.8	857	NFM_CHICK	P16053 gallus gall
10	131.5	4.8	1701	MSPI_PLAFM	P02895 plasmodium
11	131	4.8	774	GLYB_PLAFG	P02895 plasmodium
12	130.5	4.8	1183	CNA_STAAU	Q33654 staphylococ
13	130.5	4.8	1750	Y832_METJA	Q58242 methanococ
14	130	4.7	1702	IGA2_HAEIN	P45384 haemophilus
15	128.5	4.7	366	YMR3_YEAST	P33885 saccharomyc
16	128	4.7	1275	YAU9_SCHPO	Q10164 schizosacch
17	127.5	4.6	1612	TP2B_MOUSE	Q64511 mus musculu
18	127	4.6	1478	BK11_YEAST	Q01389 saccharomyc
19	127	4.6	1612	TP2B_CRILLO	Q64399 cricetus
20	126.5	4.6	1167	SCA1_STRPY	P15926 streptococ
21	126.5	4.6	1701	MSPI_PLAFB	P13813 plasmodium
22	126	4.6	515	YMR1_YEAST	P33741 saccharomyc
23	126	4.6	537	TEEG_STRPY	P18481 streptococ
24	126	4.6	682	GR78_YEAST	P16474 saccharomyc
25	126	4.6	1694	IGA0_HAEIN	P44969 haemophilus
26	125	4.6	818	PKK2_YEAST	P47116 saccharomyc
27	125	4.6	1012	SCA4_RICSL	Q9a180 rickettsia
28	123	4.5	725	ADDB_MOUSE	Q9a180 mus musculu
29	123	4.5	1076	NUPI_YEAST	Q9a180 mus musculu
30	123	4.5	1630	MSPI_PLAFK	P20676 saccharomyc
31	123	4.5	1637	MRSP_STAAU	P04932 plasmodium
32	123	4.5	1639	MSPI_PLAFW	P80544 staphylococ
33	122.5	4.5	547	CATX_BACSU	P04933 plasmodium
					P94377 bacillus su

ALIGNMENTS

RESULT 1

ID	IGA4_HAEIN	STANDARD	PRT	1849 AA
AC	P45386	AC		
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).			
GN	IGA.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBI_TaxID=727;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NHIT HK61;			
RX	MEDLINE=92234949; PubMed=1373717;			
RA	Poulsen K., Reinholdt J., Kilian M.;			
RT	"A comparative genetic study of serologically distinct Haemophilus influenzae type 1 immunoglobulin A1 proteases.";			
RL	J. Bacteriol. 174:2913-2921(1992).			
CC	!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A			
CC	PRODUCING INTACT PC AND FAB FRAGMENTS.			
CC	!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at			
CC	certain Pro- -Xaa bonds in the hinge region. No small molecule			
CC	substrates are known.			
CC	!- SUBCELLULAR LOCATION: Secreted.			
CC	!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC			
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE			
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE			
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY			
CC	SIMILARITY).			
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	M87491; AAA24968.1; ..			
DR	MEROPS; S06.001; ..			
DR	InterPro; IPR000710; Iga_S6.			
DR	InterPro; IPR004899; Pertactin_sup.			
DR	Pfam; PF023395; IGA1; 1.			
DR	Pfam; PF03212; Pertactin; 2.			
DR	PRINTS; PR00921; IGASERPTASE.			
DR	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.			
FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	1021	IMMUNOGLOBULIN A1 PROTEASE.
FT	PROPEP	1022	1849	HELPER PEPTIDE (POTENTIAL).
FT	ACT_SITE	299	299	PROBABLE.
SQ	SEQUENCE	1849 AA;	202957 MW;	79A7D018C7150AEA CRC64;

P19246 mus musculu
Q04958 saccharomyc
Q12263 saccharomyc
P58099 streptococc
P39083 saccharomyc
P75417 mycoplasma
P24482 saccharomyc
P00723 kluyveromyc
P32653 streptococc
P15205 rattus norv
Q9a183 rickettsia
P04934 plasmodium

1 NFH_MOUSE
1 YME9_YEAST
1 GIN4_YEAST
1 SCA2_STRPY
1 YGAI_YEAST
1 YD64_MYCPN
1 DPB2_YEAST
1 BGAL_KLULA
1 MRP_STRSU
1 MAPB_RAT
1 SCA4_RICAF
1 MSPI_PLAFB

34 122.5 4.5 1087
35 122.5 4.5 1679
36 122 4.4 1142
37 122 4.4 1181
38 121.5 4.4 1007
39 121 4.4 677
40 121 4.4 692
41 121 4.4 1025
42 121 4.4 1256
43 121 4.4 2459
44 120.5 4.4 1011
45 120.5 4.4 1728

```

Query Match      5.4%; Score 149; DB 1; Length 1849;
Best Local Similarity 20.1%; Pred. No. 1;
Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;

QY 28 VNKEKNAIYPHGCHADPIDEHKPVGIG-----HSHSNVELFKPEGVAKK----- 75
DB 705 LNVKGTILSGRPTHARDI-----AGISSKPKDPHTENNEVVVDDWIENPKATIM 759
QY 76 --EGN-KVYTGEELTNVALLKNSTFNQNFLLANGQKRVSPFPPELEKKGINMLVK- 131
DB 760 NVTCGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KGTDTVCVRS 802
QY 132 ----LITPDGKVLKRVSKVFGEGV-GNI-----ANFELDQYPLPGQTFKVTIASKDYP 180
DB 803 DYTYGVYTCNHSNSELKALNSFNPTNRGNVNLTENASFTLGRANLFG-----TISIGTS 857
QY 181 EVSVDGTFVTPTSLAYKMASQTIYFPHAGDTYLRVNPQFAPVKGTDALRVDFEHGNA 240
DB 858 QVN-----LKNSHHLTGNNSVNLNTNGHILHNAQNDANKVTTYNTLTVNSLSGNG 911
QY 241 -----VLENNYKVGEEKLPI-----PKLNOGT-----TRTAGNKIPVTFM 275
DB 912 SFYVWDFTNKSNKVVVKNKATCNFTLOVADKGTGEPNHELTLFDASNATRNLEVT-L 970
QY 276 ANAYLON-----QSTYIYVEVPILKENQT-----DKPSILPQFKRN 311
DB 971 ANGSDVGANKYKLRNVNGRYDLYNPEVEKRNQTVDTNITTPNDIQADAPS----- 1022
QY 312 KAQENSKLDKKEVP-----ETTEKVAENPPQENETVAKNEQEATEPTQCNCEVAKEDQP 1136
DB 340 TGNSTNSTLEEYPTDVPQOEKVAEYSGMKLENVLFNMDGTIELYLPSSGEVKKKMA 399
DB 1082 TAOQTNPNSESVPs-----ETTEKVAENPPQENETVAKNEQEATEPTQCNCEVAKEDQP 1136
QY 400 DF-----TGAPQGNENKPSNGKYSTGTVENQPT-----ENKPADSLPAPNEKPV- 447
DB 1137 TVEANTQTNATQSEKTEETO-----TATKSEPTSEVTSENQPKTVSQSTEDKVVV 1191
QY 448 ----KPNSTNDGNLNPENGVNGSDP-MLDPALDEEFAVDPVOE 485
DB 1192 EKEKAKVETEETQKAPQVTSKEPPKQAPAEVETDINAEE 1234

RESULT 2
HMW1_MYCGE
ID HMW1_MYCGE STANDARD; PRT: 1139 AA.
AC Q49413; Q49365;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome high molecular weight protein 1 (Cytochrome accessory
DE protein 1).
DE HMW1 OR MG312.
GN Mycoplasma genitalium
OS Mycoplasma genitalium
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CX NCBI_TaxID=2097;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RN SEQUENCE OF 721-847 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;

```

```

RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE (BY SIMILARITY).
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39712; AAC71534.1;
CC EMBL: U02261; RAD12527.1;
CC DR TIGR; MG312;
CC KW Cytochrome; Structural protein; Complete proteome.
CC SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;

Query Match      5.2%; Score 143.5; DB 1; Length 1139;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 128; Conservative 93; Mismatches 236; Indels 157; Gaps 33;

QY 26 IVNKEKNAIYPHGCHHH--ADPIDEH-KPVGICHSHSNVELFK-----PEEGVAKKEGN 78
DB 52 IAFNKETGVYDYPGDTYDISQLDFDENGPNFVDEKQENDYLYKRYVGNPDYSGYDNGE 111
QY 79 KYVIGEELTNVALLKNSTFNQNFLLANGQKRVSPFPPE-----LEKLGINMLVKL 132
DB 112 WWSGYFENDQWISTKESQPTDENYGFDS-----DLPEVKQSPESVEDNYGFD--ND 161
QY 133 ITPDGKVLKRVSKVFGEGVGNIANFELDPYLPQGTFKYTIASKDY----- 179
DB 162 LPPEVKQSPESVEDNY--GENDLPLPEVKQP-----ESVVDQPSDDYFAKQPTDENYGF 214
QY 180 ----PEVSY-DGTFVPTS---LAYKMASQTIYFPHAGDTYLRVNPQFAPVKGTDALVR 231
DB 215 NDLPPEVKQSPESVVDQPSDDHFAKQPESTTDSYSF---DSDL-PQPTLDQPSLDDHVQY 270
QY 232 VFD---EFHGNAYLENNYKVG--EIKLPIPKLNOQTTRTAGNKIPVTFMANAYLDNGSY 286
DB 271 NFDHHEELKPVAAEONNYOVGFDQVQANLDN-NEEIQPTAEKKVTIDF-----ESKQAO 323
QY 287 IV---EVPIL-EKENQT-----DKPSILPQKRNKAQENSKLDEKV-EPEKTSKVEK 334
DB 324 VVDSYQLPIDTDQDQDTTFSSSFETQPTV-EQFDQVNSVNDQFKPEITKEPVLESFNK 382
QY 335 EKLSTGTGNTSNTSL-----EEVPTVDPVQEKVAKFAESY- 369
DB 383 QDVVTSDLNSESNNLYSENKKNATNNDSENSEFTQLNSNSTASDDVHYESKSEPIHYK 442
QY 370 -----GMKLENVLFNMDGTIELYLPs-GEVIKKNMADFTGEAPQGNENKPS 416
DB 443 FGSDDLQSQSNNSLSESPVKFNSETAPDAHESQSEPVQVQYDI-----YQNEELKPTL 497
QY 417 NGKYSTGTVEVQPTENK-----PADSLPEAPNEKQ-----VKPENSTDMGLNPE 461
DB 498 DQPSDDYFAKQPTDENYGFNDLPPEVKQSPESVVDQPSDDHFAKQPESTTDSISDSD 557
QY 462 GNVGSDPMLDPALEAPAVDPVQ-----EKLEKFTA---SYGLGLDSVIFNMGTIELR 512
DB 558 -----LPQPTLDQPSLDDHVQYNFHDHHEELKPVAAEONNYQVGFQVQANLDNNEIQ 610
QY 513 LPSGGEVTKNLSDF 526

```



```
FT SIGNAL 1 37
FT CHAIN 38 1135
FT PROPEP 1136 1164
FT DOMAIN 434 534
FT DOMAIN 199 438
FT DOMAIN 439 526
FT DOMAIN 827 945
FT SITE 1132 1136
FT MOD_RES 1135 1135
FT SEQUENCE 1164 AA; 65DE94AF720A5474 CRC64;

Query Match
Best Local Similarity 5.1%; Score 141; DB 1; Length 1164;
Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

QY 2 KLDKKIEKI-----AGIMK--QGVKRESVWAKENAILIYPHGDHHDPI---48
DB 240 EDAEVKREELGKLFSSSTKAGLDQETQEHVKET---SEENT---QKVDEHYANSLNL 293
QY 49 -----DEHKPVGIGHSHSNYELFKPEGVAKKGNKYV--TGELTNVWNL 93
DB 294 AOKSLELDKATTNEQATQVKNQFLNACKLKEIQLIKETNVKLYKAMSESLEQVEXEL 353
QY 94 KNTFNQNTFLANGKRVYSFPEPELEKGLINMLVLTDPGKVLKVKSGVFGCVG 153
DB 354 KHNSEANLELDVAKSEIVR-----EYEGKL--NQSKNL--PELKOLEEAAHSLKQVVE 404
QY 154 NI-ANFELDQPLPGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
DB 405 DFRKKFKTSEQVTPKRVKRDLAANNENQOKIELAVSPENITVYEGEDVKFTVTA---K 460
QY 198 MASQTIFFPHAGDTYLRVNPQFVAPKGTDALRVFDEPHGNAYLENKYKGEIKLPIPK 257
DB 461 SSKT---TLDFSLLTKYNPSVS-----DRISTNYKNT---DNHKTAEITIKWLK 506
QY 258 LNOQTTRT-----AGNKIPVTFMANVLDNQSTVIVEVPILEKENQTDKPSILPQKRN 311
DB 507 LNESQVTLKAKDSSGVNVEKF-----ITVQKKEK-----QVPKT 544
QY 312 KAQENSKLDEKV-EPPKNTSEKVEKELSETGN-----STNSSTLEEVPTV-----355
DB 545 PEKDSKTEKYPQEPKSNKDKNOLOELIKSAQOQLEKLEKAIKELMEQPEIPSNPEYGIQ 604
QY 356 -----DPVQEKVAKF-----AESYGNKLENVLFNMDGTELYLPSGEVIK 395
DB 605 KSINWESQKPEIQEATSFKKIIGDSSSKYTYEYHFNKYKSDFMNQLHAQM-----EMLT 659
QY 396 KWMADFTGAPOGNGENKPSNGKYST-----GTVENQFTE-----NK-----433
DB 660 RVVQYMNKYPDPAETKIFESDMKRTKEDNYGSLNDALKGYEKYFLTPFNKIKQIVD 719
QY 434 PADSLPEAPNEKVPKPNSTNGMLNPEGNVSGDPMLOPALDEA-----PAVDPVQ 484
DB 720 DLDKVEQDQAPI-PENSE-----MDQAKKAKIAVSKYMSKVLGDGVH 762
QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
DB 763 OHLQKNNKSIKVDLFKELEAIKQOTIFDID 792

RESULT 5
SCA4_RICRH STANDARD; PRT; 1013 AA.
AC Q9AJ81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
```

```
OX NCBI_TaxID=33992;
RN [1]
RA SEQUENCE FROM N.A.
RA Sekeova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene b' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF155053; AAK30684.1;
KW Antigen.
FT NON_TER 1013 1013
FT SEQUENCE 1013 AA; 110550 MW; 856E98912315D102 CRC64;

Query Match
Best Local Similarity 5.1%; Score 140; DB 1; Length 1013;
Matches 113; Conservative 71; Mismatches 162; Indels 150; Gaps 26;

QY 48 IDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTEGELTNVWNLKNTFNQNTFLAN 107
DB 428 IDSNKQIDLKKEATAILHNAKSDIAEQ-----TNIILAEAN-TVNNQNL---473
QY 108 GQKRVYSFPEPELEKGLINMLVKLI-----TPDGKVLKVKSGVFGCVGNANFELQ 162
DB 474 -----PD-ANVAGVNAVLETIKNDQNTPD---LER--SKMLEAVAITLSENLE 517
QY 163 PYLEPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKKASQTIFFPHAGDTYLRVNPQFV 222
DB 518 PKQKEOMLEK-----TVDVGLSLK-----DDASRA---AAI 545
QY 223 PKGTDALRVFDEPHGNAYLENKYKGEIKLPIKLNQGTTRTA-GNKIPVTFMANA---278
DB 546 DGIITDVA-----IKSNLSTEDKGTMLIAGDKVKNVSELSNAEQ 584
QY 279 -YLDNQSTVIVEVPILEKENQTDKPSILPQKRNKAQENSKLD--EKVEP-----325
DB 585 KLLGSVLKKGVEAQVLSPAQO---QLMQONLDKITAEOQTKKOTIKKVNILFDPLSSTE 640
QY 326 -----PKTSEKVEKELSETGNSTNSTLE---EVPVDPVQEKVAKFA 366
DB 641 LKTTNIOAITSNVLDGPATAE-VKGEITIQITNTVAGSSLEAQDKAEIKVGVGETIATHS 699
QY 367 ESYGMKLENVLFN-----DGTIE--LYLPSGEVIKKNMADFTGEAPQNGENKPSNGKV 420
DB 700 DT-SLSLPKALIMASAEKIGIVESKTNLPBELMTKGLVDGIYE---GKGGPEITKAV 753
QY 421 STGTVENQFTEKPNPDSLPAPNEK-----PVKPNSTNGM--LNPEGVGSDPMQDPA 473
DB 754 SSG-IDNSINDSEKALKKAKDAASEATDITQNLTE-GLKQNTIEEHKPRDDIYNKA 811
QY 474 LEEAPVDPVQEKLEK 489
DB 812 QEVINAVNPVIEALEK 827

RESULT 6
TP2A_CRIGR STANDARD; PRT; 1526 AA.
AC P41515;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II, alpha isozyme (EC 5.99.1.3).
GN TOP2A OR TOP2 OR TOP-2.
```


[illegible]


```

FT REPEAT 674 596 4-2.
FT CONFLICT 216 216 S -> P (IN REF. 2).
FT CONFLICT 491 491 E -> V (IN REF. 2).
FT CONFLICT 514 514 L -> H (IN REF. 5).
SQ SEQUENCE 710 AA: 77856 MW; 9C6C8ACAE6BDF334 CRC64;

Query Match 4.8%; Score 133.; DB 1; Length 710;
Best Local Similarity 19.6%; Pred. No. 2.; Mismatches 246; Indels 232; Gaps 27;
Matches 134; Conservative 70;

QY 3 DLDKKEIEIAGIMQYGVKRESIVVNEK-----NAILYPGHGHHADPIDEHKPY---G 55
DB 64 DDDDELEPEVIDAPGVTKPRETNVPASEEIIPSTGVFPVWSDYTKP-TESVPVQEAS 122
QY 56 IGH---SHS-----NLEKPKDEGVAKKGNVYVYGEELTNVNVNLLKNSTN 99
DB 123 YGHDAFAHVRVTFSTDKBEKRDVPDIIHPLSELSDRRESRETHESLNTFVSLLSGTEDEV 182
QY 100 NQFT-----LANQKRVSVSFP-----KDFPTRSHDFDMKTIETGDNTPSRSHFEDL 346
DB 183 TSIFAPSGDDEYLDGORKVNVETPTITLEESAVSDYLSGVSNVQSKVTDPTKEETGGVPE 242
QY 120 LEKGLINMLVLIITPDGK-----VLK-----VSGKV 147
DB 243 IAESFG-NMEVTDSPDQKPGQFERDLSTRKEKEFDQDSDVLGKDSPAKPPGESGV 301
QY 148 FEGGVGNIAFELDQPLYPGQTKYTIASKDYPVSVSDGTFVTPTSLAYKMASQTIYFP 207
DB 302 FPGVGDSEGALE-----KDFPTRSHDFDMKTIETGDNTPSRSHFEDL 346
QY 208 HAGDTYLRVNPQ-FAPVKGTDALVRVDFEHGNAYLENNYKVGEIKLPIPKLNOGTRTA 266
DB 347 KTESGNDKNSPMGFGSESAGB-LEKEPDQKNDG--RNEYS-----PESDGLGLAPL 395
QY 267 GNKIPVTFEMANAYLDNSTYIVPEI-----KDFPTRSHDFDMKTIETGDNTPSRSHFEDL 346
DB 396 GGNFPYR-SHELDLKNESIDPKDVTGDPGDFLAKGRPGYGEASEEDKFPARSDDEV 454
QY 293 ---LERENQTD-----KPSILPQFKKNAQENSKLDEKVEPKTSEKVEK--EKL 337
DB 455 ETELGRDPKTEITLQFSPELSHPRERDEFKESRDDFTETRDKTEBPQKSTYTEKASML 514
QY 338 SETGNSTNSTLEEVTVD---PVQEKVAFASYNKML-----ENVLF 378
DB 515 GYSGEIPVGDQTVAGTVDKELTTPVNEKRDQETESAVTTKLPISGGSGVGEORGEDKSVS 574
QY 379 NMDGTTILYPSGEVIKKNADFTCEAPQGGNGENKPSNGKSVSTGVNENOPTENKPADSL 438
DB 575 GRDPAEAKLTITEE--DRAFSDMVAEKLIQIGBEKKE---TTTKEVEKISTE-KAASEE 628
QY 439 PEAPNEK-----PVKPE-----STDNGLNLPFEGNVGSDPML 470
DB 629 GEAVEEYKGGGMVGRIKGVFGGATDEVKPSHVEEAPKSSCGWFGGGATEEVKPKS 688
QY 471 DPALAEAP-----AVDPVQEK 487
DB 689 PHSVEESPOSLSGTVPVQEK 710

RESULT 9
NFM_CHUNK
ID NFM_CHUNK STANDARD; PRG: 857 AA.
AC PL6053;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NFM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

Db 471 ATAENAAKAEQEEQEEKAEBAEVEEBAVSEKAAEQAAAEKEEBAE-----EEBA 525
QY 364 K--FRESYMKUENLVFMDGTIELYLSGEVVKKNMADFTGEAPQNGENKPSENGKVS 421
Db 526 KSDAAEEGSKKEEIEKEEG-----EAEDEEAEAK-----GAAE 561
QY 422 TGTVENOPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMDLPALBEA---- 477
Db 562 EAGAKVEKVKSPKSPKSPKSPKSPVTEQ---AKAVQKAAAEVKGDKQAKAEKAAKEE 618
QY 478 PAVDPVQEKLEKFTA 492
Db 619 KAASPEKPKATPKVTS 633

RESULT 10
MSPL_PLAFM STANDARD: PRT: 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT Plasmodium falciparum merozoites: studies at the genetic level";
RL EMO J. 4:3823-3829(1985).
CC -/- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -/- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X05624; CAA29112.1;
DR PIR; A26868; A26868.
DR PIR; B25120; B25120.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 4.8%; Score 131.5; DB 1; Length 1701;
Best Local Similarity. 19.6%; Pred. No. 8.8; Mismatches 258; Indels 105; Gaps 20;
Matches 112; Conservative

QY 3 DLQKKEKIA-----GIMQYGVKRESIVVNVKKNAIL---YPHGDHHADPI 48
Db 435 DTKKINEKIITDNKKEKIFINIKKQIDLEKNINHTKEQNKLLDEYKSKDYELL 494
QY 49 DEHPVGIHSHSNYELFKPEEGVAKKEGNKYVTGBELTNVWLLKNSTFNNOFTLANG 108
Db 495 EKF-----YEM-KFNPNFDKVDKIFESARTYVNVKQRYNNKFSNNSSVNV 542
QY 109 Q---KRVSEFPPELEK-----KGINMLVKLITPDGVLE-KVSGKVFEGE 151
Db 543 QKLKALSYLEDYSLRKGISEKDFNHYYTLKTGLEADIKLTKTEIKSENKILEKVF-KG 601
QY 152 VGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGD 211
Db 602 LTHSANASLEVSIVKLVQVQVLLIKKIEDLRTKIELFLKNAQLKDSIHVPNIYKPNKPE 661
QY 212 TYLRVNFQFVAPKTDALVRFDEFHGNAYLNNYKVEIKLPIKLNQGTTRTAGNKIP 271
Db 662 PYLIVLKEVDKLEKFIKPKVKMLK-----KEQAVLSSTIQPLVAASE-TTEDGGHSTH 715
QY 272 V-----TFMANAVLDNOSTYIVVEPILEKENQTDKPSILPQFKRNKAQNSK- 318
Db 716 TLSQGETEVTETEVTEETVGHPTTITLPPKE-ESAPKEVKVVENSTEHKSNDNSQA 774
QY 319 LDEKVEBKTESEKVEK-----EKLSETGNSNSTLEEVPTVDPQEVKAFASYSKMLE 374
Db 775 LTKVYLKLDLFTKSYCHKYILVSNSSMDQKLLLEVNYLTPEEEKELKSCDFLDL-LF 833
QY 375 NVLENMDGTIELYLPQG-----EVIKKNMADFTGEAPQNGENKPSENGKVSQTVTIVE 426
Db 834 NQNNIPAWSYLSDMNDLQHLFFELYQEMIVYLKLEENHIKLLLEEQKQITGSS 893
QY 427 NQPTENKPADSLPEA-----PNEKPVKPESTDNGLNPEGNVSDPMDLPALBEA 481
Db 894 TSSPGNTVNTAQATHSNQSNQSNASTNTQNGV-----AVSSGPAV- 937
QY 482 PVQEKLEKFTASGLGLDSDVIFNMDGTIEL 511
Db 938 -VEESHDPDLTVL-----SINDLKGVSL 960

RESULT 11
GLYB_PLAFG STANDARD: PRT: 774 AA.
ID GLYB_PLAFG
AC P02895;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycophorin-binding protein precursor (GBP-130).
GN GBP.
OS Plasmodium falciparum (isolate FCR-3 / Gambia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5838;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86133561; PubMed=3512098;
RA Kochan J., Perkins M., Ravetch J.V.;

RT "A tandemly repeated sequence determines the binding domain for an
 RT erythrocyte receptor binding protein of P. falciparum.";
 RL Cell 44:689-696(1986).
 RN [2]
 RP SEQUENCE OF 556-774 FROM N.A.
 RX MEDLINE=85142172; PubMed=3883491;
 RA Ravetch J.V., Kochan J., Perkins M.;
 RT "Isolation of the gene for a glycoprotein-binding protein implicated
 RT in erythrocyte invasion by a malaria parasite.";
 RL Science 227:1593-1597(1985).
 CC -!- SOURCE: CYTOPLASM OF SCHIZONTS. RELEASED
 CC INTO THE CULTURE SUPERNATANT AT THE TIME OF SCHIZONTS RUPTURE.
 CC ONLY A MINOR FRACTION OF GBP WAS FOUND TO BE LOOSELY ASSOCIATED
 CC WITH MEROZOITES.
 CC -!- DEVELOPMENTAL STAGE: SYNTHESIZED AT THE TROPHOZOITE AND SCHIZON
 CC STAGES.
 CC -!- SIMILARITY: TO PLASMODIUM GLYCOPHORIN BINDING PROTEIN-RELATED
 CC ANTIGEN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M12897; AAA29608.1; -;
 DR EMBL; M10985; AAA29607.1; -;
 DR PIR; A03390; ZOZQMF;
 DR PIR; A24057; A24057;
 DR InterPro; IPR003681; GBP_repeat;
 DR Pfam; PF02526; GBP_repeat; 11;
 KW Merozoite; Malaria; Repeat; Signal.
 FT SIGNAL 1 69
 FT CHAIN 70 774 GLYCOPHORIN-BINDING PROTEIN.
 FT DOMAIN 226 774 11 X 50 AA TANDEM REPEATS.
 FT REPEAT 226 275
 FT REPEAT 276 325
 FT REPEAT 326 375
 FT REPEAT 376 425
 FT REPEAT 426 474
 FT REPEAT 475 524
 FT REPEAT 525 574
 FT REPEAT 575 624
 FT REPEAT 625 674
 FT REPEAT 675 724
 FT REPEAT 725 774
 FT CONFLICT 556 556
 FT CONFLICT 563 563
 FT CONFLICT 570 570
 FT CONFLICT 640 640
 FT CONFLICT 690 690
 FT CONFLICT 747 747
 FT CONFLICT 774 774
 SQ SEQUENCE 774 AA; 90017 MW; 16B795D08C0C1798 CRC64;
 Query Match 4.8%; Score 131; DB 1; Length 774;
 Best Local Similarity 21.1%; Fred. No. 3.1;
 Matches 92; Conservative 60; Mismatches 161; Indels 124; Gaps 20;
 QY 140 LKVSQKVFEGVGNTANFELDPQYLPQGTQKYT---TASQDYPEVSYDGTFTVPTSLAY 196
 Db 3 LSKVS-DIKSTGVSNKFN-----SNSSKYSLEMSVSKNEKNSLGAFHSKILLIF 55
 QY 197 KVASQTIFPFHAGDTYLRV-----NPFAPVDPKGDALVRVDFEFGHAYLENNYKVGSEI 251
 Db 56 GIIVYVLLNAY-ICGDKYKAVDYGFRSRIILAEGETCAR-----KE 97
 QY 252 KLPVPLNGGT-TRTAG-----NKIPVTETMANAYLDNQ---STIV--EVPLEKE 296
 Db 98 KTLKRSKOKTSTRVAOTKKADEKNSVVEEQVSDSEKQKTKKVKVKKQINIGOTE 157
 QY 297 NOTDKPSILPQFKRKAQENSKLDE-----KVPEPKTSEKVEKEKLSGTNS---- 343

Db 158 NOKGKNNKVKIKKKEKKEGKPEENKHNANEASKKPRASKVSKPSTSTRSNNEVKI 217
 QY 344 TSNSTLEEVPTVDVQEKVAKFA--ESYGMKLE---NVLFNMDGTIELYLPSEGVIKK-- 396
 Db 218 RAASNQETLTSADPEGOIMREYAADPEYRKHLEIFYKILINTD-----PNDEVERNA 270
 QY 397 -NMADFTGEAPQNGENKPSENG-----KVSIGTVENQPTENKPADSLPEAPNEK 445
 Db 271 DNKEDLTASDEPGOIMREYASDPEYRKHLEIFYKILINTDNDDEVERNAD----- 321
 QY 446 PVKPNSTNDGMLNPEGNVSGDPMLEAPADVPQVEKLEKFTASYGLGDSVIFNM 505
 Db 322 -----NKEDLTASDEGOIMREYAADP-----EYRKHLEVF-----HKILINT 359
 QY 506 DGTIELRLPSGEVKKN 522
 Db 360 D-----PNDEVERN 369

RESULT 12

CNA_STAAU STANDARD; PRT; 1183 AA.
 ID CNA_STAAU AC Q53654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen adhesin precursor.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FDA 574;
 RX MEDLINE=92165839; PubMed=1311320;
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
 RA Lindberg M., Hoeoek M.;
 RT "Molecular characterization and expression of a gene encoding a
 RT Staphylococcus aureus collagen adhesin.";
 RL J. Biol. Chem. 267:4766-4772(1992).
 RN [2]
 RP ERRATUM.
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
 RA Lindberg M., Hoeoek M.;
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC STRAIN=FDA 574;
 RX MEDLINE=94032261; PubMed=8218209;
 RA Patti J.M., Boles J.O., Hoeoek M.;
 RT "Identification and biochemical characterization of the ligand
 RT binding domain of the collagen adhesin from Staphylococcus aureus.";
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
 RX MEDLINE=97475225; PubMed=9334749;
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,
 RA Narayana S.V.L.;
 RT "Structure of the collagen-binding domain from a Staphylococcus
 RT aureus adhesin.";
 RL Nat. Struct. Biol. 4:833-838(1997).
 CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
 CC COLLAGEN-CONTAINING SUBSTRATA.
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M81736; AAA20874.1; .
 CC PDB; 1AMX; 24-JUN-98.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR TIGR; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 1154
 FT PROPEP 1155 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 FT SITE 1151 1155
 FT MOD_RES 1154 1154
 FT SEQUENCE 1183 AA; 133066 MW; B6A1CC072E57D76 CRC64;
 Query Match 4.8%; Score 130.5; DB 1; Length 1183;
 Best Local Similarity 21.6%; Pred. No. 6;
 Matches 99; Conservative 62; Mismatches 197; Indels 101; Gaps 19;
 QY 74 KEGKV-VTGEELTNVLLKNSTFNQFTLANGQKRVSFSPPELEKLGINKLVKL 132
 Db 680 KARGQVKTVEELTKV-----KGYTHVDNNDM--GNLIVTKYTPETTSISGEKVWDK 733
 QY 133 ITPDGKVLKSVKVEG-----VGNIANFELDQVLP---GQTFKYTIAS---KD 178
 Db 734 DNQGRKPEKSVNLLADGEKVKTLDTSETNKKYEFKDLPRYDEGKKIETVIVTETHVKD 793
 QY 179 YPEVSDGCTFTVPTSLAYKMSQTFYPHAGDTLVY-----NQFAVPKGTDAIVRV 232
 Db 794 Y-TDINGT-----TIINKYTPGETSATVTKNNDNNQDGKRPTEIKVEL 838
 QY 233 FDE-----FHGNAYLENNKYGEIKLPIPKLNQ---GTRTAGNKKIPVT 273
 Db 839 YQDKATGKATILNESNNHTHTWGLDEKAKQGVKYTYVEELTKVGYTHVDNNDMGNL 898
 QY 274 FVANAYLDNOSTVIVPEILEKNQDK-----PSILQPKRNKA-----Q 314
 Db 899 IVTKYTPETTSISGEKVWDKDNQDGKPEKSVNLLANGEKVKTLDTSETNKKYEFK 958
 QY 315 ENSKLDE--KVEPKTSKVEKEKLSCTNSNSFL--EEVPTVD-----P 357
 Db 959 DLPKYDEGKKIETVIVTETHVKDYTTDINGTITINKYTPGETSATVTKNNDNNQDGKRP 1018
 QY 358 VQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFT--GEAPQNGENKPS 416
 Db 1019 TEIKVELYQDGKATGKATILNESNNHTHTWGLDEKAKQGVKYTYVEELTKVNGYTHVD 1078
 QY 417 NGKVSCTVENQTEKNKPADSL-PEAPNEK--PVKPSNS 452
 Db 1079 NNDMGNLIVTKYTPKKNPKNPIPEKPKDKTPTTPKPDHS 1117

RESULT 13
 Y832_METJA
 ID Y832_METJA STANDARD; PRT; 1750 AA.
 AC Q58242;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0832 [Contains: Mja rnr-1 intein; Mja rnr-2 intein].
 DE MJ0832.
 GN Methanococcus jannaschii.
 OS Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;

CC SEQUENCE FROM N.A.
 RN STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RC MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glöck A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- PFM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: WEAK IN THE C-TERMINAL, TO M.JANNASCHII MJ0885.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; U67527; AAB98834.1;
 DR HSP; P17255; IDFA.
 DR TIGR; MJ0832;
 DR InterPro; IPR005144; ATP.
 DR InterPro; IPR003586; Hedgehog_hintC.
 DR InterPro; IPR003587; Hedgehog_hintN.
 DR InterPro; IPR002203; Intein.
 DR InterPro; IPR004042; Intein_endonuc.
 DR Pfam; PF03477; ATP_cone; 2.
 DR PRINTS; PR00379; INTEIN.
 DR SMART; SM00305; HincC; 2.
 DR SMART; SM00306; HincN; 2.
 DR PROSITE; PS00818; INTEIN_C_TER; 2.
 DR PROSITE; PS00819; INTEIN_ENDONUCLEASE; 2.
 DR PROSITE; PS00817; INTEIN_N_TER; 2.
 KW Hypothetical protein; Autocatalytic cleavage; Protein splicing;
 KW Complete proteome.
 FT CHAIN 1 337
 FT CHAIN 338 790
 FT CHAIN 791 1058
 FT CHAIN 1059 1591
 FT CHAIN 1592 1750
 FT SEQUENCE 1750 AA; 203279 MW; 6D68938770A25359 CRC64;
 Query Match 4.8%; Score 130.5; DB 1; Length 1750;
 Best Local Similarity 20.2%; Pred. No. 10;
 Matches 124; Conservative 113; Mismatches 225; Indels 151; Gaps 36;
 QY 1 MKDLK----KIEKIAIGIMKQYGVKRESIVYNNKEKNAIYPHGDHHDPIDEHKPVGI 56
 Db 347 IKEGDKLVKIGEAIDFEMEKY---KDKIIVDGDTEIL-----YLDGIAEVYISV 395
 QY 57 GHSNRYELFKPEGVAK--KEGNKVY--TGELTNVLLKNSTFNQFTLANGQKRV 112
 Db 396 NVKTKGAK--FKRYVAISRKHPRGVKVGKDGTSIIIVTETHDHSFLNFYDENGLNVCVAPRQ 454
 QY 113 -----SESPPELEKLGINKLVKLITPDGK-----VLEKSGKVFGEVGNIAF 159.
 Db 455 MKHIIRNENFYDVEYRIGDYIETNYQRTDSKYNQRNDIPEKL--KITKELCQLGLFV 512
 QY 160 LDQVLPQGTQKTYTASDKYPEVSDGTFT---VPTSLAYKMSQTI-----FYFP--- 207

```
Db 513 AGSVI---TNGISITTKDDDDIAKIEFVKEQINENTAVKRYEDSVRFVNGYRFLKE 569
Qy 208 HAGDTYLRVN-PQFAPVPGTDAALVRV--DEFHGNAYLENNYKV-----GEIKLP 254
Db 570 HINGRAINKNSEF-ILKGDREMKLAFGLGISDGYVSKDGRVQIYVTSQQLGLQLHL 628
Qy 255 IFKLQ-----GTRTAGNKIPVFWANAYLDNOSTYIVVEVILEKENOTD-KPSILPOF 308
Db 629 LSDGLMIVSITKIEEGKIEI--KRNEIVRNKYLVIEI-----AKNCTEDLKYVIEKY 682
Qy 309 KKNKAQ-----ENSKLDERVEEPTKEKVEKLSGTSTGN-----STLBEVPTVDV-V 358
Db 683 KKERIKPANYDOLPYDRIKLEHLAKITDKKPYNDYANKSNRKLKLTLEKIEQLNPHL 742
Qy 359 QKVKAFKESYGMKLENVLFNMDGTIELYLPSEGVIKNMADEFTGEAPQGNENKPSNG 418
Db 743 REEINKF-----KL-NIPFEIKEKEI-----DNGYVYDLSVED--NENE 780
Qy 419 KYSTGTVENQPTENKPADS-----LPAPNEKPKVKPENSTDNM----- 457
Db 781 ITATGIL-----CHNIFISINLEIPEFLDKKPAVIAGTTRGYDVEERAKLILEALV 836
Qy 458 -LNPNGVSGDPMPLD-----ALEEAPADVPQOE-----KLEKFTASVGLDLSVFNMDGTI 509
Db 837 DVWMEGDAMGKPFLEPNFTIKLRNFAFKDENKELMYKIHQLSAKFSI---PVFINN--- 889
Qy 510 ELRLPSGEVIKKN 522
Db 890 ---LPDWQVNTN 899

RESULT 14
ID IGA2_HAEIN STANDARD; PRT: 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulson K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases.";
CC Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT EC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
```

```
DR EMBL; M87489; AAA24966.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF02395; IGAI.1.
DR Pfam; PF03212; Pertactin.2.
DR PRINTS; PRO0921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT REPEAT 1109 1116 K.
FT REPEAT 1117 1124 2.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 4.7%; Score 130; DB 1; Length 1702;
Best Local Similarity 20.5%; Pred No.11;
Matches 106; Conservative 61; Mismatches 201; Indels 150; Gaps 20;

Qy 72 VAKKEGKVV-----YTGEELTNVYNLLKSTENNQNFTLANGOKRVSFSPFPELEKGLI 126
Db 783 IGYKAGDTVCVRSYDTG-YVTCTTDLKSLDKALNSFNATVSGVNLNLS-----GNANFVLGK 837
Qy 127 NMLYKLITPDG----KVLKVSQKVGCEGVGNANFELDQPYLPQGTFKYTIASKD---Y 179
Db 838 ANLFGTISGTGNSQVRULTENSHHLLTGDSNVNQLNDKGHIHLNAQDANKVTYNTLTV 897
Qy 180 PEVSYDGTFTVPTSLAYKMASQTTFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEPHGN 239
Db 898 NSLSGNGSFYLLDLSNKGQDKVVTGATGNFTLQVADKTGEP--TKNELTLFDA--SN 953
Qy 240 AYLENNYKVGKILPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIVVEPPILEKENQT 299
Db 954 A-TRNNLNVSLV-----GNTVDLGAMKYKLRNVNRYDLYNPEVEKRNQT 997
Qy 300 -----DKPSI-----LPQKRNKAQNSKLDER--- 322
Db 998 VDTNITPNNIQADVPSVPSNNEEIARVETPVPPAPATPSETTETVAENSQESATVE 1057
Qy 323 -----VEEPTKSEK--VEKEKISETGNSTNSLTLEVPVDPVQ--EKVA 363
Db 1058 KNEQDATETTAQNGEVAEEAKPSVKANTQTNEVAQSGSETEETQTTEIKETAKVEKEKA 1117
Qy 364 KFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGNENKP----- 414
Db 1118 KVEKEEKAKVEK-----DEIOEPQOMASETSFKQAKKAPKE 1153
Qy 415 -SENGKYSTGTVENQ-----TENKPADSL--PEAPNEKPKVPK---ENSTD 454
Db 1154 VSTDTKVEETQVQAQPTQSTTVAABEATSPNSKPABETQPSKTNAPETPVVSVKQOTE 1213
Qy 455 NGMLNPEGNVSGDPMPLDPALEEAFAV---DPVQEKLE 488
Db 1214 NTDQPTEREKTAKEVTEKTEQPPQVQASQAPKQEQSE 1251

RESULT 15
YNR3_YEAST
ID YNR3_YEAST STANDARD; PRT: 366 AA.
AC P33885;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 40.3 kDa protein in RPS3-PSD1 intergenic region.
GN YNL173C OR N1673.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
```


OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 23.447 Seconds
(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKEETAGIMKQYG.....TELRLPGEVKKNLSDIFA 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database :

- SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteria:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2734	99.6	1039	16 Q9ANY1	Q9any1 streptococc
2	184	6.7	2004	16 Q37QP7	Q97qp7 streptococc
3	169.5	6.2	1964	2 Q59947	Q59947 streptococc
4	155	5.6	1524	10 Q8RYN2	Q8ryn2 oryza sativ
5	153	5.6	1873	2 Q924N7	Q924n7 enterococcu
6	149.5	5.4	4888	16 Q8PQ08	Q8pq08 ureaplasma
7	148	5.4	1939	5 Q25662	Q25662 plasmodium
8	147	5.4	2151	5 Q9NG79	Q9ng79 trichomonas
9	146.5	5.3	791	3 Q9USW1	Q9usw1 schizosacch
10	146	5.3	1272	16 Q98PR8	Q98pr8 mycoplasma
11	145.5	5.3	1078	5 Q963T1	Q963t1 plasmodium
12	145.5	5.3	1927	2 Q54875	Q54875 streptococc
13	142	5.2	2485	5 Q96134	Q96134 plasmodium
14	141.5	5.2	1441	16 Q9CFL1	Q9cfl1 lactococcu
15	141	5.1	1134	2 Q99051	Q99051 streptococc
16	140.5	5.1	17352	5 Q95YM2	Q95ym2 procambarus

17	140	5.1	841	3 Q94002	Q94002 candida alb
18	139.5	5.1	2443	5 Q9VSA2	Q9vsa2 drosophila
19	139	5.1	750	16 Q9PP44	Q9pp44 campylobact
20	138.5	5.0	1127	5 Q94248	Q94248 caenorhabdi
21	137.5	5.0	387	16 Q97KL1	Q97kl1 clostridium
22	137.5	5.0	1377	3 Q42695	Q42695 candida alb
23	137.5	5.0	1526	11 Q55079	Q55079 cricetulus
24	137.5	5.0	2768	5 Q9VC00	Q9vc00 drosophila
25	137	5.0	940	16 Q928N7	Q928n7 listeria in
26	136.5	5.0	940	16 Q8Y4N9	Q8y4n9 listeria mo
27	135.5	4.9	881	2 Q9RFJ4	Q9rfj4 streptococc
28	135	4.9	825	2 Q93GT5	Q93gt5 streptococc
29	135	4.9	1255	2 Q982J6	Q982j6 shigella so
30	134.5	4.9	1049	5 Q9VI49	Q9vt49 drosophila
31	134.5	4.9	1764	2 Q93T34	Q93t34 haemophilus
32	134.5	4.9	2025	11 Q99PP2	Q99pp2 mus musculu
33	134.5	4.9	2481	16 Q99QR6	Q99qr6 staphylococ
34	134	4.9	380	11 Q63575	Q63575 rattus norv
35	134	4.9	2276	2 Q93TV6	Q93tv6 staphylococ
36	133	4.8	710	10 Q39060	Q39060 arabidopsis
37	133	4.8	1020	5 Q27104	Q27104 trichomonas
38	133	4.8	5458	5 Q90459	Q90459 plasmodium
39	132.5	4.8	806	16 Q9CKX5	Q9ckx5 pasteurella
40	132.5	4.8	1037	13 Q73808	Q73808 fugu rubrip
41	132.5	4.8	1526	11 Q55078	Q55078 cricetulus
42	132	4.8	1231	2 P72362	P72362 streptococc
43	132	4.8	1653	2 Q9LBC3	Q9lbg3 streptococc
44	132	4.8	2478	2 Q9LCH2	Q9lch2 staphylococ
45	132	4.8	2478	2 Q9RL69	Q9rl69 staphylococ

ALIGNMENTS

RESULT 1

Q9ANY1 PRELIMINARY; PRT: 1039 AA.
ID AC Q9ANY1
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004).
GN PHE OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langemann S., Koenig S., Johnson S.;
RA "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis."
RL Infect. Immun. 69:949-958(2001).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson A.S., Heidelberg J., DeBoy R.F., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RL pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AF318956; AAK06761.1;

DR EMBL; AE007403; AAK75121.1; -
DR TIGR; SP1004; -
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 29 POTENTIAL
SQ SEQUENCE 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match 99.6%; Score 2734; DB 16; Length 1039;
Best Local Similarity 99.6%; Pred. No. 1.3e-135;
Matches 526; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKDLKKIEKIAIGKQYKRESIVVAKENAIYIPGHGHHADPIDEHKPVGIGHSH 60
Db 512 MKDLKKIEKIAIGKQYKRESIVVAKENAIYIPGHGHHADPIDEHKPVGIGHSH 571

QY 61 SNYELFKPEEGVAKKGNKYVTGEELTNVNLKNTFTNQNFTLANGOKRVSFPEPPEL 120
Db 572 SNYELFKPEEGVAKKGNKYVTGEELTNVNLKNTFTNQNFTLANGOKRVSFPEPPEL 631

QY 121 EKKLGINMLVKLITPDGKVLKESYGVFGEVGNIANFELDPYLPQGTFKYTIASKDYP 180
Db 632 EKKLGINMLVKLITPDGKVLKESYGVFGEVGNIANFELDPYLPQGTFKYTIASKDYP 691

QY 181 EYSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
Db 692 EYSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 751

QY 241 YLENNYKVGKIKLPIPKLNGTTRTAGNKTPVTFMANAYLDNOSTYIVVEVPILEKENQTD 300
Db 752 YLENNYKVGKIKLPIPKLNGTTRTAGNKTPVTFMANAYLDNOSTYIVVEVPILEKENQTD 811

QY 301 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSTLEEVTVPDVPQVE 360
Db 812 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSTLEEVTVPDVPQVE 871

QY 361 KVAFASGYGKMLNVLNMDGTIELYLPSEVYIKKNMADFTGEAPQNGENKPSSENGKV 420
Db 872 KVAFASGYGKMLNVLNMDGTIELYLPSEVYIKKNMADFTGEAPQNGENKPSSENGKV 931

QY 421 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPNPEGVSDPMLDPALEAPAV 480
Db 932 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPNPEGVSDPMLDPALEAPAV 991

QY 481 DPVQEKLEKTASYGLDGLSDVIFNMMDGTIELRLPSEVYIKKNLSDPTA 528
Db 992 DPVQEKLEKTASYGLDGLSDVIFNMMDGTIELRLPSEVYIKKNLSDPTA 1039

RESULT 2
Q97QP7 PRELIMINARY; PRT; 2004 AA.
ID Q97QP7
AC Q97QP7; PubMed-11453916;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
GN SP1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTG84;
RX MEDLINE=21357209; PubMed-11453916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Mayhew L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguilo S., Dickson T., Hickey E.K.,
RA Holt J.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
FT "Complete genome sequence of a virulent isolate of Streptococcus

pneumoniae".
RT Science 293:498-506(2001).
RL EMBL; AE007416; AAK75263.1; -
DR MEROPS; M26.001; -
DR TIGR; SP1154; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001130; Zn_MTpeptidse.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; TSIRK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004 AA; 223908 MW; 556BC6A1028D60A4 CRC64;

Query Match 6.7%; Score 184; DB 16; Length 2004;
Best Local Similarity 23.5%; Pred. No. 0.27; Mismatches 192; Gaps 34;
Matches 130; Conservative 70;

QY 60 HSNYELFKPEEGVAKKGNKYVTGEELTNVNVN-----LLKNSTFN-----NQN 102
Db 279 HKNLET-KKEKISPKKTKGNTLNPNQDEVLSGOLNKPPELLYREETMETKIDFOEEIQEN 337

QY 103 FTLANGOKRVSFPEPPELKKLGINM-LVKLIT-----PDGKVLKES 144
Db 338 PDLAGTTRV-----KOEGLKGGKVEIVRIFSVNKEVSEIVSTSTAPSPRIVEKGT 391

QY 145 GK--VFGE-----GV-----GNIANFELDPYLPQGTFKYTIASKDYPSYDGTFTVP 191
Db 392 KKTQVIKQSPETGVHVKDQSSAIVEPAL-QPELP-----EAVSDKGEVQV---TLP 442

QY 192 TSLAYKMASQTIYFPFHAGDTYLRVNPQ----FAVPKGTDALVRVDFEHGNAVLENNYK 247
Db 443 EAVV-----TDKGET--EVQPESPDTVSDKGEPEQVAPLPYKGN-----481

QY 248 VGEIK-LPIPKL-NOGTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENOTDKPSI 304
Db 482 IEQVAPETPEVTEKQGEPEKT--EEVPV-----KPTETPNNEGTEGTSI 527

QY 305 LPQFRN---KAQENSKLDEKVEEPTSEK-----VEKEKLETSNGS 343
Db 528 --QEAENPVQPAEESTNTSEKV-SPDTSKNTGEVSSNPSTTSVSGSNKPEHNSKNE 584

QY 344 TSNSTLEEVTVPDVPQEKVAKFAESYGMKLENVLNMDGTIELYLPSEVYIKKNMADFTG 403
Db 585 NSEKTVVEVP-VNP-----NEGTVE-----GTSNGET 610

QY 404 EAPQNGENKPSSENGKV---STGTVENQPTENKP--ASLPEAPNEKPKVSPENS---TD 454
Db 611 EKPVPQPAEETQTNQSGKIANENTGEVSNKPSDSKPPVEESNQPEKNGATKPKNSGNTTSE 670

QY 455 NGMLAPE---GNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYGLDGLSDVIFNMMDGTIEL 511
Db 671 NGQTPPEPSNGNSTDEVSTESNTSNGNEEKQENE-----LDPKKKVEEPEKTELEL 723

QY 512 RLPSGEVYIKKNLSD 525
Db 724 R-----NVSD 728

RESULT 3
Q59947 PRELIMINARY; PRT; 1964 AA.
ID Q59947
AC Q59947;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoglobulin A1 protease (EC 3.4.24.13).
GN IGA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=R6;
RC  MEDLINE=97047667; PubMed=8926056;
RA  Wani J.H., Gilbert J., Plaut A., Weiser J.N.;
RT  "Identification, cloning and sequencing of the Immunoglobulin A1
RL  protease gene of Streptococcus pneumoniae.";
RL  Infect. Immun. 64:3967-3974(1996).
DR  EMBL; U47687; AAC44568.1; -
DR  MEROFS; M26.001; -
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  InterPro; IPR00130; Zn_Mtpeptidse.
DR  Pfam; PF00746; Gram_pos_anchor.1.
DR  TIGRFAMs; TIGR01167; LPXTG_anchor.1.
DR  TIGRFAMs; TIGR01168; YSRK_signal.1.
DR  PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR  PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW  Hydrolase; Protease.
SQ  SEQUENCE 1964 AA; 218786 MW; 0599E818A783B7E8 CRC64;

Query Match          6.2%; Score 169.5; DB 2; Length 1964;
Best Local Similarity 23.7%; Pred. No.1.5;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY  60 HSNYLFKPEEGVAKKEGNKVT-----GEELTNVYN-----LLKNSTFN-----NQN 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  279 HKNLET-KKEEXISPKKGTGVNLNPQDEVLSQINKEPILLYREETIETKIDFQEEIGEN 337

QY  103 FTLANGOKHVSFPPELEKLGINM-LVKLIT-----PDGKVLKVS 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  338 PDLAAGTVRV-----KQEGKLGKVEIVRIFSVNKEEVSREIVSTSTAPSPRIVEKGT 391

QY  145 GK--VFGE-----GV-----GNIANFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVP 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  392 KKTQVKEQEPETGVHEKDVQSGAIVEPAI-QPELP-----EAVVSKGPEVQP-----TLP 442

QY  192 TSLAYKMASQTFYPPHAGDIYLRNPQ-----FAVPKGTDALVRVDFDFHGNAYLENYK 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  443 EAVV-----TDKGET--EYQPSPTVWSDKGPEQVAPLPKYGN-----481

QY  248 VGEIK--LPIPKL-NOGTRTAGNKIPYTFMANAYLDNQSTIVIVPILKRENOTDKPSI 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  482 IQQVRPEPVEKTKQGGKPT--EEVPV-----RPTETPNVNEGTEGTSI 527

QY  305 LPQFKN-----RAQENSKLDEKVEPKTEKV-----EKKLSETGNST 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  528 --QEAENPVQPAEESTTNSEKVSPTDTSSENTEGVSSENFSDTTSVGSNKPKPHNDSKNEN 585

QY  345 SNSTLEEYTPDPVQEKVAKFAESYGMKLENVFNMDGTIELYLPSPGEVIKKNAFTGE 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  586 SKTVEEVP-VNP-----NEGIVE-----GTSNQETE 611

QY  405 APQGNKPKSPNGKV---STGTVENOPTENKP--ADSLPEAPNEKPKVPKPS-----TDN 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  612 KPQPAEETQTSNGKIANENTGEVSNKPSDSKPPVEESNQPEKNGKATKPKPSNGTITSEN 671

QY  456 GMLNPE 461
Db  672 GQTEPE 677

RESULT 4
QRYN2
ID  Q8RYN2 PRELIMINARY; PRT; 1524 AA.
AC  Q8RYN2;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE  P0563E10.13 protein.
GN  P0563E10.13
OS  Oryza sativa (japonica cultivar-group).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=39947;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV_NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL  clone:P0563E10.";
RL  Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AP004317; BAB90730.1; -
SQ  SEQUENCE 1524 AA; 168701 MW; E47EE1A9C5842228 CRC64;

Query Match          5.6%; Score 155; DB 10; Length 1524;
Best Local Similarity 19.5%; Pred. No.6.2;
Matches 107; Conservative 77; Mismatches 206; Indels 160; Gaps 22;

QY  9 EKIA--GIMKOYGVKRESIVNKR--EKNATYIPGHGHHADPIDEHKPVGIGHSHSN-- 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  498 DEKLTMRIRVKEETYEETALVNKNKGSSAVNLTHCDHNEKGVLPQPKPQHTAQSGSKLE 557

QY  63 -----YELFKPEEGVAKKEGNKVTGEBELTNVNLKNSTFNNONFTLANGOKR 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  558 QLGKWTSGAEFYVLISPDQ---KCKTNSV-----TCEGDNVQTTNPSSK 598

QY  112 VSFSPPELEKLGINMLKITPDGKVLKVSQKVFEGV---GNIANFELDQPLPG 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  599 LG-----QFEKKGK-----ETTSQDFVGCCKSWDGGDIAELRHEHVNLR- 637

QY  168 QTFKYTIASKD-----YPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQF 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  638 --EYAGTSTEDGCKAPTAPETSPN--EKPT---YQESTETHFKECVGAQNY-----682

QY  221 AVPKGTDALVRVDFDFHGNAYLENYKVEIKLPIPLKNOGTTTA-----GNKTP-- 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  583 -----QERTGDDGAFEISCDSSKLAHPGASLESCISGCHCNKNSPSD 729

QY  272 -----VTFMANAYLDNQSTIVIVPILKRENOTDKPSILPQFKNKAQEN-----SKL 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  730 ASTRETTSLGESKNENNALEVPQAD-EMQSOILQYHEFRNENIDENKASQVKVSKL 788

QY  320 DEKVEEKP-----SEKVEKEL-----STGNSTSNSTLEEYTPDPVQEKV 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  789 EESVEYETPNFQKSSSTAHTGETETVEKEMFSFDELFPQKNKNGITTEAPPESLIHKEI 848

QY  363 AKFA--ESYGMKLENVFNMDGTIE-----LYLPSPGEVIKKNM 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  849 KKFCTEEKAYITLGGDVVQKSGSLERANITLSASANEENEEAENAFVEGINVETHV 908

QY  399 ADFTGEAPQNGENKPKSPNGKVSTGTVENOPTENKPADSLPEAPNEKPKVPKPSNDGML 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  909 STY-GTSVEDSDQIQSENRMMDGMDLVSHGNEEAADPWLNDSEKSSQVEEIFSHEEGQL 967

QY  459 NPEGNVGSDP 468
Db  968 SVEGGIDGGP 977

RESULT 5
Q9Z4N7
ID  Q9Z4N7 PRELIMINARY; PRT; 1873 AA.
AC  Q9Z4N7;
DT  01-MAY-1999 (TrEMBLrel. 10, Created)
DT  01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Surface protein precursor.
OS  Enterococcus faecalis (Streptococcus faecalis).
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC  Enterococcaceae; Enterococcus.
OX  NCBI_TaxID=1351;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MMH594;
RX  MEDLINE=99081742; PubMed=9864215;

```

RA Shankar V., Baghdayan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
 RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
 RL a gene encoding a novel surface protein.";
 DR Infect Immun 67:193-200(1999).
 DR EMBL: AF034779; RAD09858.1; -;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRfam: TIGR01167; LPXTG_anchor; 1.
 DR TIGRfam: TIGR01168; YSRK_signal; 1.
 KW Signal.
 FT SIGNAL 1 49 POTENTIAL.
 FT CHAIN 50 1873 SURFACE PROTEIN.
 FT SEQUENCE 1873 AA; 202084 MW; F609483DB9B80621 CRC64;
 Query Match 5.6%; Score: 153; DB 2; Length 1873;
 Best Local Similarity 23.7%; Pred. No. 10;
 Matches 123; Conservative 58; Mismatches 194; Indels 144; Gaps 28;
 QY 2 KLDKKI---EKGAGIMKQGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIG 57
 DB 1027 KDLFKDAKVTWKEVD--ISTAGNKKGTVVVT-----YSGSDEVE-VD---VTVT 1072
 QY 58 HSHSNLEFKPE--EGVAKKEGKNVVTGELTNVNVLLKNSTENN-----CNFTLANGQ 109
 DB 1073 DNRSDADKYEPTVEGKEVIGGKVDLTQVNLPTLPQGTITVDTPGTTIDTNPNGYE 1132
 QY 110 KRVSFSPPELEKKGILNMLVKL---ITPDGKVLKSVKSGVFGEGV-----GNIAN 157
 DB 1133 GVIEVTPD-----GTKDVTVPVVEVTQNRSDADKYEPTVEGKEVIGGKVDLTQVNL 1186
 QY 158 FELDQVLPQGTQFKYTIASKDPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVN 217
 DB 1187 L-----PTLPQGT---TVT-----DVTGPGTIDTNPNGYEVIEV-----TY---- 1221
 QY 218 POPAVPKGTDALVRVDFEFGHAYLENNY-----KV---GEIKLP-----IPKLNQGT 262
 DB 1222 -----PDGKDTVKVPVEVTDNRSDADKYEPTVEGKEVIGGKVDLTQVNLPTLPQGT 1276
 QY 263 T---RTAGNKIPVTFMANAYLDNQSY-----IVEPTILEKENOTDKPSILPQFKRKA 313
 DB 1277 TVDVTGPGTIDTNPNGYEVIEVTPDGTQDVKVPVEVTDNRSDADKYEPTVWGEKV 1336
 QY 314 QENSKLDEKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKL 373
 DB 1337 EIGGKVD-----LTDNVNLPTLPQGTITVDTPG-----GTID 1370
 QY 374 ENVLNMDGTIELYLPSPG--EYTK-----KNMADTGEAPQNGENKPESENGKYS-TGT 424
 DB 1371 TNPNGYEVIEVTPDGTQDVKVPVEVTDNRSDADKYEPTVEGE-KVEIGGKVDLTQV 1429
 QY 425 VENOPTENKPADSLPEAPNEKPKYPENSTNDNGMLNPEGN 463
 DB 1430 VTNLPT-----LPQGTITVDTPGTTIDT---NTPGN 1458
 RESULT 6
 Q9PQ08 PRELIMINARY; PRT: 4688 AA.
 ID Q9PQ08
 AC Q9P008;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein U0482.
 GN U0482.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OC NCBI_TaxID=134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RL urealyticum.";
 DR Nature 407:757-762(2000).
 DR EMBL: AE002145; AAF30894.1; -;
 DR InterPro: IPR001152; Thymosin_b4.
 DR SMART: SM00152; THI; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEEL1997E CRC64;
 Query Match 5.4%; Score 149.5; DB 16; Length 4688;
 Best Local Similarity 20.6%; Pred. No. 54;
 Matches 125; Conservative 89; Mismatches 219; Indels 173; Gaps 27;
 QY 1 MKDLDDKIEKIAIGIMKQGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
 DB 3859 INDQQTINVTLSGVNSKYN-GROIKVYKDNVNIY-----ESSLITLQKG 3905
 QY 61 SNYELFPEEGVAKKEGKNVVTGELTNVNVLLKNSTFNNQNTLANGQKRVSFSPPEL 120
 DB 3906 NDYQLL-----LSNLNSNREYRFEKIE--INHISN-TNNEDELEKLG---VSNFTITQT 3954
 QY 121 EKKL-----GINMLVKLITPDGKVLKSVKSGVFGEGVGNIANFELDQYLP 166
 DB 3955 KNTTVQWNSSATIVGTRGVNFNFKIKSED-KILEN-----NQGVVAMFAP 3999
 QY 167 QTEK-----YTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVN 219
 DB 4000 KETIRDNTWLTQYTRPLKDVTSDFEGTWAHDLNSNVNFKET-----TYKLKVIQ 4050
 QY 220 FAVPKGTDALVRVDFEFGHAYLENNY-----KVGEIKL-PIPKLNQGTITRT--- 265
 DB 4051 F-VNKPTKAKNNI--NSENNAVILDNTNSINSNYEFTTKVGDHKLINITSSNNVNTNSQTI 4108
 QY 266 -----AGNKIPVTFMANAYLDNQSYIVEPTILEKENOTDKPSILPQFKRKAQ 314
 DB 4109 NFTLSGVKKSQVWKKIKLSYKSN-----DTSESHHTNEVLIESNKTQYNILLNKKRNTY 4164
 QY 315 ENSKLEKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKL- 373
 DB 4165 -TLIDYKLDNNVNSDPPREGNLTNSFIITRTSAINVLEIEISNASTNLKSTIIKIN 4222
 QY 374 -----ENVLNMDGTIELYLPSPGSEVVKKNMADTGEAPQNGENKPESENGKYSTGT----- 424
 DB 4223 LNDPDLNRDQDQATIVY-----GNNKQAMGFITVSGNIKLYLTATLVLDN 4267
 QY 425 -----VENQPTENKPA-----DSLPEAPNEKPKYPENSTND---NGM 457
 DB 4268 FNDKNTVNIISFNKKPSTAAENIGDKSNIIYNDSIP-----KLEINNDIIVNGP 4319
 QY 458 LNPE-----GNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLD-----SVIFNMD 506
 DB 4320 INKELIVKANQKNNDVLDGLQINPKIAHNLRFIAKFKSTNNDIETNVIINGSSLYVND 4379
 QY 507 GTIELR 512
 DB 4380 GKTSIR 4385
 RESULT 7
 Q25662 PRELIMINARY; PRT: 1939 AA.
 ID Q25662
 AC Q25662;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Repeat organellar protein.
 OS Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5825;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=96V;

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA McDougall R.C., Rajandream M.A., Barrell B.G., Skelton J.,
 RA Churcher C.M.,
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL121794; CAB57927.1;
 KW Hypothetical protein.
 SQ SEQUENCE 791 AA; 85523 MW; 02579D93CF5D68B8 CRC64;

Query Match 5.3%; Score 146.5; DB 3; Length 791;
 Best Local Similarity 22.1%; Pred. No. 7.1;
 Matches 115; Conservative 67; Mismatches 183; Indels 155; Gaps 26;

QY 46 DPIDEHKPY-GIGHSHSYELFKPEGVAKEGKVTGTELTNYYVLLKSTFNQNT 104
 DB 126 DPEDAGVPLAGL-----EESTDNVEWQDFATNEKLFV-----KSHFEDLYT 168
 QY 105 LANGQKRVSEFPPELEKLLGINMLVK---LITPDGKVLKVKSGVFGEV-----G 153
 DB 169 -----SRIDRSHPKYKEKEQEDRAKEIEGIVTNIHIAERGLKVDGSLDEEDLYSG 223
 QY 154 NIANFELDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTY 213
 DB 224 VHRSIDVVRNYT--RSNAYKNKKQKPKNHEAPHQP-----QKVVVP---DDPA 270
 QY 214 LRNVPOFANPK-GTD--ALRVVFDEFHGNAYLENNYKVEIK-----LPIKLNQ 261
 DB 271 IVSHRLALPRAPGPDPSRAAEFFNARRKAGPLSRREKDGQKEFMQFSQSLKIGSLD-- 328
 QY 262 TTRTAGNKIPVTFMANAYLDNSTIVIVEPILEKENQNDKPSILPQFRNKAQENSKLDE 321
 DB 329 -----SKQP-----SSTKSAEVKVADEKQLDPASS-----QATPADSKPRKE 367
 QY 322 KVEEPT-----SEKVEKEKLGSTNSTSTLEEVTPVDVQEK-VAKFAESYGMKL- 373
 DB 368 EAEKPYTSATEVSERKEV-----VDGNTSPSKEEKEPSTEPEKPSVVTQRKETTGLG 423
 QY 374 -----ENVLFNDGTTLELPSGEVILKKNADFTGEAPQ----- 407
 DB 424 TKLNKAIEKPNVAPVTPGKFTIPSPAPVNASRPMQPSNNSASIPSTTPQSPS 483
 QY 408 --GNGENKPESE-----NGKYST-----GTVENQPTENKPAD-----SLPEAPNEK 445
 DB 484 VVSGENKPSPPVFPFNGPVSSKEKEPILDNFNVKNGVBEHQGAEQIDKPFSCPTWNTG 543
 QY 446 PVKPNSTNGMLN--PEGNVGSDPMLDPALEAPAVDPV 483
 DB 544 P-----LSLOQTANSRPEGNSGS-----AKKAAANPM 572

RESULT 10
 Q98PR8 PRELIMINARY; PRT; 1272 AA.
 ID Q98PR8
 AC Q98PR8;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hypothetical protein MYPV_6510.
 GN MYPV_6510
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267155; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Callison F.,

RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis"; 2145-2153(2001).
 RL Nucleic Acids Res. 29:2145-2153(2001).
 DR EMBL; AL445565; CAC13824.1;
 DR MYPULIST; MYPV_6510;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1272 AA; 147169 MW; 90019B854E6C6D74 CRC64;

Query Match 5.3%; Score 146; DB 16; Length 1272;
 Best Local Similarity 20.3%; Pred. No. 14;
 Matches 125; Conservative 94; Mismatches 238; Indels 158; Gaps 28;

QY 1 MKQDDKIEEKIAGIMKQYGVKRESIVVKNKNAIIVPHGDHHDHDPIDHKPVGIGHSH 60
 DB 86 LTFDITQDDEKINVSYK-----LISNEQK-----QEENSORNSNQ 121
 QY 61 SNYELFPPEGVAKKEGKVTGTELTNYY-----NLLKNSTFNQNTLANGQKRVSE 114
 DB 122 QNEKFANDDEGTLKVLVEFFSYNGVPVTRVYKLGKFKNTIQKTSNNLNLNLSNKEQ----- 177
 QY 115 SFPEL-----EKLGINMLVK---LITPDGKVLKVKSGVFGEVGNIANFELDQPLPG 168
 DB 178 IHPSELDSQQVLTTFELKSLLETLDEKYVLYKVVYVNDIDAEAKIRLTQLKTNE 237
 QY 169 TFK--YTIASKDYPEVSYDGT-----TVPTSLAYKMASQTIFFPHAGDTYLRNP-- 218
 DB 238 NIKNSFEVLKDKFKYKLNDDSSLKAKINLVKNLSTIEELNITSKGFANKTEDDVNKLK 297
 QY 219 --QFAPKGTDALRVDFEFHGNAYLENNYKVEIK--KLPIKLNQGTTRTAGNIPVT 273
 DB 298 ETLNIDLPNGYEF-----EFVSLAPKANDASVGLLTLYKLVKNLNGTSENPSK-- 347
 QY 274 FMANAYLDNSTIVIVEPI--ILEKENQNDKPSILPQFRNKAQENSKLDEKVEEPTSEK 331
 DB 348 --NGRIET-NTIVEETNLNLSKNDKDEP-----KDNSTNSKDEPKTDEKVEEP 397
 QY 332 VEKE-KLSEGTGNTSNTLEE-----VPTVDVQEKVA----- 363
 DB 398 REDEPKTNMPSKDKPKIDEPNEPKDQKTEPKNEPKPKPKVEPADETLAIPDKISKI 457
 QY 364 -----KFAESYGMKLEN--VLFNMDGT--IELYLPSEGEVILKKNADFTGEA 405
 DB 458 ELKENSQKLPQSPKES-DLNLNLSKLVSDCKNKFSELSLPQGYSTFKLA----- 510
 QY 406 PQNGENKPSNGKVSQTV-----ENQPTENKPADSLPEAPNEKPKPENSTNGMLN 459
 DB 511 -----SNSNDEGTLVKVIVQKQKVKTKELKLTLLTEFESLK-----ESDFQDLSN 561
 QY 460 PEGNVGSDPMLDPALEAPAVDPVQEK-----LEKFTASYGL-GLDSVFNNMDGTIELR 512
 DB 562 KKDRLASVVNDKIIKESLV--VKNTIENFDENKIDISYSSSLDEV---NGKLKIK 615
 QY 513 LPSGEVILKKNLSDFI 527
 DB 616 MTIEKTKDRLKEEL 630

RESULT 11
 Q963T1 PRELIMINARY; PRT; 1078 AA.
 ID Q963T1
 AC Q963T1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Glutamate-rich protein (fragment).
 GN GLURP.
 OS Plasmodium reichenowi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5854;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=21313688; PubMed=11420113;
RA Thelsen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RT glutamate-rich protein (GLURP) from Plasmodium reichenowi.";
RL Mol. Biochem. Parasitol. 115:269-273(2001).
FT EMBL; AF356828; AAK40236.1;
FT NON_TER 1
FT NON_TER 1078
FT SEQUENCE 1078 AA; 123581 MW; 80388088B223913C CRC64;
SQ
Query Match 5.3%; Score 145.5; DB 5; Length 1078;
Best Local Similarity 20.0%; Pred. No. 12;
Matches 111; Conservative 74; Mismatches 176; Indels 195; Gaps 23;
QY 2 KDL-----DKLEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDH--HHADPIDEHKPKVGI 56
DB 161 KQLQDLDLEPPPTQIHKDY--KEENLI--NEEDSEPPRHEHKKLDNHYEKNVHF 216
QY 57 GHS--HSNYELFKPE-----GVAKE-----GNKYTGELTNVNVNLLKNSTFN 99
DB 217 ENSVNSNOENLKPSPDEHLNINSKLEDOLDLHEHGNSEHLKDE-----KGNPELV 270
QY 100 NONFTLANGQKRVSPFPPELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFE 159
DB 271 HENLSPNDPIEQIILNLEPQ-----ETNVQQLYNEKQNVEEKPNSSEIFSLDLKETTND 325
QY 160 LDOPYLPGOTKTYITASKDYPEVSYDGTFTVPTSILAYKMASOTIFYPPHAGDTVLRVNPQ 219
DB 326 I----LPNQ-----NPL 333
QY 220 FAVPGKTDALVRVDFEFGHGNAYLNNYKVGKIPKLNQGTTRTAGNKIPVTFMANAY 279
DB 334 ENIKQSESEINHVDH-----ALPEN-----IIDK 359
QY 280 LDNSTYIVE---VPILEKENQDKPSILPQKRN--KAQENSKLDEKVEBPKTSERKEK 334
DB 360 LDNEKEHIDESHQNVNVLQ--ENNINNHOLEPQERANIESFEPKIDSEIILP-----ENVK 415
QY 335 EKL-----SETGNSTNSTLEPYTPDVOEKVAKFAESYGMKLENVLF 378
DB 416 BEIYVDVPSPKHFHNETLEQETSESEHEAVSEKNAHETVHEEAVSQESNPERADN-- 472
QY 379 NWDGTIELYLSGSEVINKNMAFTGCEAPQNGENKPSNGKVSCTGTVENQPTENKPADSL 438
DB 473 --DGNV-----SONSNELNENEFVESEKSEHEFAENES--SL 507
QY 439 PEAPNEKPKPENSTDNG-----MLNPEGNVSDPMLDPALEAPAVDFVQEKLEKETASY 494
DB 508 EEAHQEEIVPQNQNGESGSKLVNDGE-----GFEAHEEIVPQNQNGES--- 554
QY 495 GLGLDSVIFNWDGTIE 510
DB 555 --GSKLVNDNDEGGFE 568
RESULT 12
Q54875 PRELIMINARY; PRT; 1927 AA.
ID Q54875
AC Q54875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Iga1 protease.
GN IGA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK81.
RX MEDLINE=97047666; PubMed=8926055;
RA Poulsen K., Reinholdt J., Kilian M.;

```

```

RT "Characterization of the Streptococcus pneumoniae immunoglobulin A1
RT protease gene (iga) and its translation product.";
RL Infect. Immun. 64:3957-3966(1996).
DR EMBL; X94909; CAA64396.1;
DR MEROPS; M26.001;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMS; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1927 AA; 215170 MW; BAF74DB811301037 CRC64;
Query Match 5.3%; Score 145.5; DB 2; Length 1927;
Best Local Similarity 22.4%; Pred. No. 27;
Matches 126; Conservative 79; Mismatches 207; Indels 151; Gaps 34;
QY 3 DLDKKIEETAGIMQYGVKRESIVVNKKNALIIYPHGDH--HHADPIDEHKPK--VGIGH 58
DB 197 DFEVSNQEKSAAT-----PTKQKQVDVNVTFNFDHSTVQAIQEQTPVSTKPTQVVE 252
QY 59 SHSNYELFKPEEGVAKKQKVTGELTNVNVNLLKNSTFNQNTLANGQKRVSPFP 118
DB 253 KPFSTELINPR-----KEEKSSDSQEQLAHEKMLETKK-----EEKIS--PK 293
QY 119 ELKKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDPYLPQGTFTYITASKD 178
DB 294 E---KTGVNLT---NPQDEVLS-----GOLNK----- 314
QY 179 YPEVSYDGTFTVPTSILAYKMASOTIFYPPHAGDTVLRVNPQFAVPKGTDALVRVDFEFG 238
DB 315 -PELLYREE-TIEKIDFQEEIQE--NPDLAECT--VRKQEGKLGKRVK--IVRIFS---- 364
QY 239 NAYLENNYKVG-----EIKLPIKL-NQGTTRTAGNK-IPVTFMANAYLDNSTYIVEY 290
DB 365 ----VNKEEVSREIVSTSTTAPSPRIVEKTKTKVIKEQPETGV--EHKDVQSGAIVE- 417
QY 291 PILEKE-----NOTDKPSTLPQFKRNKAQENSK--LDEKVEBPKTSEKVEKEKLETSGN 342
DB 418 PAIQPELPEAVVSKGEPEVQPTLPFAVVTDKGEPAVQPELPEAVVSKGEPEVQAPLPE 477
QY 343 STNSLTSEETVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEV--IKKNMADF 401
DB 478 YKGN--IEQVKPETVE---KTEQGEKPEEVPVK-----PTEETPVNPEGIT 522
QY 402 TGEAPQNGEN--KPSNGKVSCTGTVENQPT--ENKPADSLPEA-----PNEKPVKP 449
DB 523 EGTSLQG--AENPVQAEEDTQINSKIANENTGEVSNKPSDSKPPVEESNQPKNGTATKP 581
QY 450 ENS-----TDNGMLNPE---GNVSDPMLDPALEAPAVDFVQEKLEKFTASYGLGDSVI 502
DB 582 ENSGNTTSENGQTEPEPSNGNSTEDVSTKNTSNSNGNEIKQENE-----LDPDKKV 634
QY 503 FNMGTIELPLSPGEVIKKNLSD 525
DB 635 EDPEKLTLELR-----NVSD 648
RESULT 13
O96134 PRELIMINARY; PRT; 2485 AA.
ID O96134
AC O96134;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ser/Thr protein kinase.
GN PFB0150C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]

```

	SEQUENCE FROM N.A.	
RPR	MEDLINE=99021743; PubMed=9804551;	
GARA	Gardner M.J., Tetellin H., Carucci D.J., Cummings L.M., Aravind L.,	
XRA	Koonin E.V., Shalom S., Mason T., Yu K., Fujii C., Pederson J.,	
RRA	Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,	
SARA	Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,	
TARA	Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;	
PART	"Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.";	
RT	Science 282;1126-1132(1998).	
LRL	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.	
CCC	EMLL; AE001376; AAC71820.1;	
DNR	InterPro: IPR000719; Euk_pkinase.	
DRR	InterPro: IPR002290; Ser_thr_pkinase.	
DDR	Pfam: PF00069; pkinase; 1.	
DDR	proDom: PD000001; Euk_pkinase; 1.	
DDR	PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.	
DDR	PROSITE: PS00108; PROTEIN_KINASE_ST; 1.	
KKW	ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.	
WSQ	SEQUENCE 2485 AA; 293764 MW; 4297C5CF0F03AD7E CRC64;	
	Query Match 5.2%; Score 142; DB 5; Length 2485;	
	Best Local Similarity 20.0%; Pred. No. 57;	
	Matches 110; Conservative 64; Mismatches 171; Indels 204; Gaps 25;	
QY	2 KDLDKLTDEKTAGIMKQGVK-----22	
	::: : ::: :	
DDb	1078 KEMEKKKMEKEKVMEKEMEVEKEVRELKNMNMNNMMNNEKMNEINYKNELIY 1137	
QY	23 ----RESIVNVKNEKNAIIYPGHGHHADPIDERK---PVGIGHSHSNY----- 63	
DDb	1138 VDNDKELEIVEEREK-LIIPFNYES-----DVHKNMMSININCKDDNYNLKEYVDNS 1191	
QY	64 -----ELFPESGVAKKE-----GNKYT---GEELTVNVLLKNSTFNQNQT- 104	
DDb	1192 CIAQKEENIFRPFLNALNKDKVKWRFKNIKIITIHNEMKRIYTINKNVFFPIYNFR 1251	
QY	105 ----LANGOKRVSTSFPPLEKKLGINLVKLIPTDGKVLKVSQVFGEVGN--IANF 158	
DDb	1252 YENFLIN--HLTYNFK-----NDLFKL-----SYKVMNNIRNLYIA NK 1289	
QY	159 ELDQQY----LPQG---TKFYITASXDYPE---VSVDGTFTVPTSLAYKMASQTIFYPF 207	
DDb	1290 HINNHYDYMNKLQNVIYTLKIQVANIDNDHHTCKGGGLDYTNMNISCEKNR----- 1343	
QY	208 HAGDTYLVRNQFAVPKGTDALAVDFEPHGAYL-----ENNYSVGGEIKLPIPK 257	
DDb	1344 -KDKITL--NKIFHYKKKDKARFINDEICSDMYDIKKKYSNDENNYKLN----- 1393	
QY	258 LNQGTRTAGNIKIPVTFMANAYLDNQSYTIVEVPILLEKENQTDKPILPOFQRNKAQENS 317	
DDb	1394 -----KMNISMSNDEDMT---PTLNSEHGHNFPSCQPNLLLEKK---ST 1430	
QY	318 KLDEKVEBPKTSKVEKEKLSETGNSNSTLEEVPITDPQEVKAFAESYGKMLENVL 377	
DDb	1431 YIDLNLDSNSMDPDTEKYNFVNEN-----DLFNTKRKFNFSGKLNFNKK 1479	
QY	378 F----NMGDTIELPSGEVIKKNHADFTG-----EAPOGNGENKPSGENKYSTG 423	
DDb	1480 FFVSVNEDGVTFSS-----KNNLFRNELKSNLSLKLESVKNSNNCSNKKGDGDNIG 1531	
QY	424 TVENQPTEN 432	
DDb	1532 NMENMNTTN 1540	
	RESULT 14	
QCFLl	PRELIMINARY; PRT: 1441 AA.	
ID	Q9CFL1	
AC	Q9CFL1	
DC	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:52 : Search time 8.18433 seconds
(without alignments)
1898.176 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2146

Sequence: 1 MKDLKKIEBKIAIGIMKQYG.....IELRPSGEVIRKNSLDFIA 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184	6.7	571	4	US-08-961-083-4
2	169.5	6.2	1964	2	US-08-790-912-3
3	169.5	6.2	2052	2	US-08-790-912-2
4	151	5.5	278	4	US-08-961-083-94
5	149	5.4	1848	4	US-08-296-791-6
6	149	5.4	1848	5	PCT-US95-10661A-6
7	143	5.2	1164	4	US-08-923-992A-10
8	141	5.1	1098	4	US-08-923-992A-8
9	141	5.1	1164	4	US-08-923-992A-2
10	140	5.1	1128	4	US-08-923-992A-6
11	139	5.1	1104	4	US-08-923-992A-4
12	138.5	5.0	984	1	US-08-242-932-2
13	138.5	5.0	984	1	US-08-714-481-2
14	138.5	5.0	984	5	PCT-US95-08111-2
15	132.5	4.8	1164	2	US-08-589-756-1
16	132.5	4.8	1164	4	US-09-206-800-1
17	132.5	4.8	1164	4	US-09-206-898-1
18	132	4.8	1231	3	US-08-904-263A-4
19	130.5	4.8	1183	2	US-08-447-031A-2
20	130	4.7	1702	4	US-08-296-791-5
21	130	4.7	1702	5	PCT-US95-10661A-5
22	129	4.7	10182	4	US-09-134-001C-3159
23	128.5	4.7	933	3	US-08-293-728-2
24	128.5	4.7	933	4	US-09-421-868-2
25	126.5	4.6	932	4	US-09-071-035-416
26	126.5	4.6	969	4	US-09-071-035-414
27	126.5	4.6	1167	2	US-08-589-756-2

28 126.5 4.6 1167 4 US-09-206-800-2 Sequence 2, Appli
29 126.5 4.6 1167 4 US-09-206-898-2 Sequence 2, Appli
30 126 4.6 682 1 US-08-441-139-2 Sequence 2, Appli
31 125.5 4.6 345 4 US-08-856-253-7 Sequence 7, Appli
32 125 4.6 1183 4 US-09-134-001C-3530 Sequence 3530, Ap
33 122.5 4.5 3696 4 US-09-134-001C-5080 Sequence 5080, Ap
34 122 4.4 1181 4 US-09-206-898-23 Sequence 23, Appli
35 121 4.4 763 4 US-08-961-083-66 Sequence 66, Appli
36 121 4.4 982 4 US-09-071-035-246 Sequence 246, App
37 121 4.4 962 4 US-09-071-035-250 Sequence 250, App
38 121 4.4 962 4 US-09-071-035-254 Sequence 254, App
39 121 4.4 962 4 US-09-071-035-470 Sequence 470, App
40 121 4.4 962 4 US-09-071-035-474 Sequence 474, App
41 121 4.4 962 4 US-09-071-035-478 Sequence 478, App
42 118.5 4.3 1545 4 US-08-296-791-4 Sequence 4, Appli
43 118.5 4.3 1545 5 PCT-US95-10661A-4 Sequence 4, Appli
44 118 4.3 1125 4 US-09-513-783A-152 Sequence 152, App
45 118 4.3 1610 4 US-09-513-783A-22 Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-961-083-4
; Sequence 4, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-4

Query Match 6.7%; Score 184; DB 4; Length 571;
Best Local Similarity 23.5%; Pred. No. 6e-07;
Matches 130; Conservative 70; Mismatches 162; Indels 192; Gaps 34;

QY 60 HSNVELFKPEGVAKKEGKNVYT----GELTNVNV-----LLKNSTFN-----NQN 102
DB 119 HKNLET-KREKISPKETKTGVNTLPQDEVLSGLINKPELLYREETMETKIDFQEEIQEN 177

103 FTLANGOKRVSEFPPELEKKGINM-LVKLIT-----PDGKVLKVS 144
178 PDLAEGTVRV-----KQEGKLGKKEIVFISVKNKEVSREIVSTSTAPSRIVEKGT 231
145 GK--VFGE-----GNIANFELDQYLPQGTFKYTIASKDYPEVSDGTFTVP 191
232 KKTQVIKEQPTGVHKGVSQGAIVEPAI-QPELP-----EAVSDKGEPEVQ 282
192 TSLAYKMASOTIYFPFHAGDTYLRVNPQ-----FAVPKGTDLALVRVDFHGNAYLENNYK 247
283 EAVV-----TDKGET--EVQPSPTVSDKGEPEQVAPLPEYKGN----- 321
248 VGEIK--LPIPKL-NOGTRTAGNKPVTETMANAYLDNQSTIYVEVPILEKENOTDKPSI 304
322 IQQVKEPTVPEKTEKQGEPT--EEVPV-----KPTETVPNNEGTEGTSI 367
305 LPQKRN---KAQENSKLDEKVEEPTSEK-----VEKEKLTSEGNS 343
368 --QEAENPVQPAEESTTNSKRV-SPDTSKNTGEVSSNPSTSTSVGESNKPENHDSKNE 424
344 TSNSTLEEVPYDVPQVKVAKFAESYGMKLENVLFNMDGTIELVLPSEVIVKKNMADFTG 403
425 NSEKVEEVP-VNP-----NEGTV-----GTSNOET 450
404 EAPQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPVKPKENS---TD 454
451 EKPQPAEETQNSKIANENTGEVSNKPSDSKPPVEESNQPKNGTKATKPKENSGNTTSE 510
455 NGMLNPE---GNVSDPMDLPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIEL 511
511 NGQTEPEPNSGNTSEVSTESNNGNEEKQENE-----LDPKKVKEEPEKTELEL 563
512 RLPSGEVIVKKNLSD 525
564 R-----NVSD 568

RESULT 2
US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020

TELEFAX: (215) 567-2891
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 6.2%; Score 169.5; DB 2; Length 1964;
Best Local Similarity 23.7%; Pred. No. 5.9e-05;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY 60 HSNVYELFKPEEGVAKKEGNKYVT---GEELTNVNV---LLKNSTFN-----NON 102
DB 279 HKNLET-KKEEKISPKETGNTLNPDQDEVLSGOLNKPPELLYREETIETKIDQBEEIOEN 337
QY 103 FTLANGOKRVSEFPPELEKKGINM-LVKLIT-----PDGKVLKVS 144
DB 338 PDLAEGTVRV-----KQEGKLGKKEIVFISVKNKEVSREIVSTSTAPSRIVEKGT 391
QY 145 GK--VFGE-----GNIANFELDQYLPQGTFKYTIASKDYPEVSDGTFTVP 191
DB 392 KKTQVIKEQPTGVHKGVSQGAIVEPAI-QPELP-----EAVSDKGEPEVQ 442
QY 192 TSLAYKMASOTIYFPFHAGDTYLRVNPQ-----FAVPKGTDLALVRVDFHGNAYLENNYK 247
DB 443 EAVV-----TDKGET--EVQPSPTVSDKGEPEQVAPLPEYKGN----- 481
QY 248 VGEIK--LPIPKL-NOGTRTAGNKPVTETMANAYLDNQSTIYVEVPILEKENOTDKPSI 304
DB 482 IEQVKEPTVPEKTEKQGEPT--EEVPV-----KPTETVPNNEGTEGTSI 527
QY 305 LPQKRN---KAQENSKLDEKVEEPTSEK-----VEKEKLTSEGNS 344
DB 528 --QEAENPVQPAEESTTNSKRV-SPDTSKNTGEVSSNPSTSTSVGESNKPENHDSKNE 585
QY 345 TSNSTLEEVPYDVPQVKVAKFAESYGMKLENVLFNMDGTIELVLPSEVIVKKNMADFTGE 404
DB 586 SEKVEEVP-VNP-----NEGTV-----GTSNOET 611
QY 405 APOQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPVKPKENS---TDN 455
DB 612 KPVQPAEETQNSKIANENTGEVSNKPSDSKPPVEESNQPKNGTKATKPKENSGNTTSEN 671
QY 456 GMLNPE 461
DB 672 GQTEPE 677

RESULT 3
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-2

Query Match          6.2%; Score 169.5; DB 2; Length 2052;
Best Local Similarity 23.7%; Pred. No. 6.3e-05;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY 60 HSNVLFKPSGVAKKGNKYVT---GEELTNVNV---LLKNSTN-----NQN 102
DB 352 HKNLET-KKEKISKPKETGNTLNPDQVLSGOLNPELLYREETIETKIDFOEEIQEN 410
QY 103 FTLANGOKRVSFSPPELEKKLGIM-LVKLIT-----PDGKVLKVS 144
DB 411 PDLAEGTVRV-----KQEGKLGKVKVILRVFSVNKEVSRVISTTAPSPRIVEKGT 464
QY 145 GK--VFGE-----GNIANFELDQYLPFGQTFKYTIASKDYPEVSYDGTFTVP 191
DB 465 KKTQVIKEQPTGVHEKDVQSGAIVEPAI-QPELP-----EAVSDKGEPEVQP---TLP 515
QY 192 TSLAYKMASQIFVFFHAGDYLRVNVQ----FAVPKGTDALVRVDFEFGNAYLNNYK 247
DB 516 EAVV-----TDKGET--EVQPSPTVVSVDKGEQVAPLFEYKGN-----554
QY 248 VGEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILEKENOTDKPSI 304
DB 555 IEQVKPETPVPEKTRQEGPEKT--EEVPV-----KPTETPVNPNNEGTEGTSI 600
QY 305 LPQPKRN---KAQENSKLDEKVEEPKTSKV-----EKEKLSGTENST 344
DB 601 --QEAENPVQPAEESTTNSEKVSPTDTSSENTGEVSSNPDSSTTSVGSNKPHEHNDKSEN 658
QY 345 SNSTLEEYPTDVPQVEKFAESYGMKLENVLFNMDGTIELYLPSSGEVYKKNWADFTGE 404
DB 659 SEKTVEEVP-VNP-----NEGTV-----GTSNQETE 684
QY 405 APQNGENKPKSENGKV---STGTVENOPTENKP--ASLPEAPNPKPVKPKNS-----TDN 455
DB 685 KPVQPAEETQNSKIANENTGEVSNKPSDSKPPVERSNQPEKNGTATKPNNSGNTTSEN 744
QY 456 GMLNPE 461
DB 745 GQTEPE 750
```

```

RESULT 4
US-08-961-083-94
; Sequence 94, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Chol et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-94

Query Match          5.5%; Score 151; DB 4; Length 278;
Best Local Similarity 23.7%; Pred. No. 0.00011;
Matches 77; Conservative 44; Mismatches 102; Indels 102; Gaps 17;

QY 224 KGTDALVRVDFEFGNAYLNNYKVGKIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYL 280
DB 19 KGEPEQVAPLPEYKGN-----IEQVRPETVEKTKEGQPEKT--EEVPV-----60
QY 281 DNGSTVIVEPILEKENQTDKPSILPQKRN---KAQENSKLDEKVEEPKTSK-----331
DB 61 ---KPTETPVNPNNEGTEGTSI--QEAENPVQPAEESTTNSEKV-SPDTSSKNTGEVS 113
QY 332 -----VEKELSETGNSNSTLEEVYPTDVPQVEKFAESYGMKLENVLFN 379
DB 114 SNPSDSTTSVGSNKPHEHNDKSENSEKTVEEVP-VNP-----150
QY 380 MDGTIELYLPSSGEVYKKNWADFTGEAPQNGENKPKSENGKV---STGTVENOPTENKP--434
DB 151 NEGTV-----GTSNQETEKPVPQPAEETQNSKIANENTGEVSNKPSDSKPPV 199
QY 435 ADSLPEAPNPKPVKPKNS-----TDNMLNPE---GNVGSDDMLDPALEAPAVDPVQEKL 487
DB 200 EESNQPEKNGTATKPNNSGNTTSENGQTEPEPSNGNSTEDVSTESNTSNGNEEIKQEN 259
QY 488 EKTASVGLDLSVIFNMDGTIELR 512
DB 260 E-----LDPDKKVEPEKTLLELR 277
```

```

RESULT 5
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
```

us-09-471-255-10.ra1

Tue May 13 14:29:39 2003

; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791
 ; FILING DATE: 25-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1389
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1848 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; US-08-296-791-6

Query Match 5.4%; Score 149; DB 4; Length 1848;
 Best Local Similarity 20.1%; Pred. No. 0.0026;
 Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
 QY 28 VNKEKNAIYPGHDHHPADIDEKHPVGIG-----HSHSNVLEFKPEGVAKK----- 75
 DB 704 LNVEKGLFLSGRTPPHARDI-----AGISSTKDPHTENNVEVDDWLNRFKATM 758
 QY 76 --EGN-KVYTGEELTNVNLKSTFNQNTLANGOKRVSFSPPELEKKGIMLVK- 131
 DB 759 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KTGDTVCVRS 801
 QY 132 -----LITPDGKLVKSVGKVEGEGV-GNI-----ANFELDQYPLPGQTKYTIASKDYP 180
 DB 802 DYTGYVTCHNSNLSEKALNSFNPTNLRGNVNLTENASFTLGKANLFG-----TIQSIGTS 856
 QY 181 EYSYDGTFTVPTSLAYKMASQTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
 DB 857 QVN-----LKNESHWHLTGNSNVNQLNLTHGHLHNAQNDANKVTTNTLTNLSLNG 910
 QY 241 -----YLENNYKVGEEKLPI-----PKLNOQT-----TRTAGNKIPVTM 275
 DB 911 SFYVWVDFTNKSNKVVVKNKSNATGNFTLQVADKGTGEPNHNELTLDASNATRNLEVT-L 969
 QY 276 ANAYLDN-----QSTYIVVEPILEKENOT-----DKPSILPQFKRN 311
 DB 970 ANGSVDRGAWKYLKRVNNGRYDLNPEVEKRNQTVDTNITTPNDIQADAPS----- 1021
 QY 312 KAGENSKLEKVEEP-----RTSEKVEKEKISE 339
 DB 1022 -AQSNNNEELARVETPPPPAPATASIASEQPETPAETAQAPAMEETNTANSTETAPKSD 1080
 QY 340 TGNSTNSLTLEEVPTVDPVOEKVAKFAESYGMKLENLVNMDGTTIELYLPSCGEVKKWA 399
 DB 1081 TATQTNPNSESVPs-----ETTEKVAENPPQENETVAKNEQATEPTPQNGEYAKEDQP 1135
 QY 400 DF-----TGEAPOGNENKPSNGKYSTGTVENOPT-----ENKPADSLPEAPNEKPV- 447
 DB 1136 TVEANTQTNEATSEGTETETQ-----TAEKTEPTSESTVTSNENPEKTVSSTEDKVVV 1190
 QY 448 ----KPENSTDNGLNPEGNVSGDP-MLDFALFEAPVDPVOE 485

DB 1191 EKEERAKVETEETQAPQVTSKPPKQAEPAPEEVPVTDINADE 1233
 RESULT 6
 PCT-US95-10661A-6
 ; Sequence 6, Application PC/TUS9510661A
 ; GENERAL INFORMATION:
 ; APPLICANT: Washington University, et al.
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/10661A
 ; FILING DATE: 16-AUG-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/296,791
 ; FILING DATE: 25-AUG-1994
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: FP-59941/RFT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1389
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1848 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; PCT-US95-10661A-6

Query Match 5.4%; Score 149; DB 5; Length 1848;
 Best Local Similarity 20.1%; Pred. No. 0.0026;
 Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
 QY 28 VNKEKNAIYPGHDHHPADIDEKHPVGIG-----HSHSNVLEFKPEGVAKK----- 75
 DB 704 LNVEKGLFLSGRTPPHARDI-----AGISSTKDPHTENNVEVDDWLNRFKATM 758
 QY 76 --EGN-KVYTGEELTNVNLKSTFNQNTLANGOKRVSFSPPELEKKGIMLVK- 131
 DB 759 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KTGDTVCVRS 801
 QY 132 -----LITPDGKLVKSVGKVEGEGV-GNI-----ANFELDQYPLPGQTKYTIASKDYP 180
 DB 802 DYTGYVTCHNSNLSEKALNSFNPTNLRGNVNLTENASFTLGKANLFG-----TIQSIGTS 856
 QY 181 EYSYDGTFTVPTSLAYKMASQTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
 DB 857 QVN-----LKNESHWHLTGNSNVNQLNLTHGHLHNAQNDANKVTTNTLTNLSLNG 910
 QY 241 -----YLENNYKVGEEKLPI-----PKLNOQT-----TRTAGNKIPVTM 275
 DB 911 SFYVWVDFTNKSNKVVVKNKSNATGNFTLQVADKGTGEPNHNELTLDASNATRNLEVT-L 969
 QY 276 ANAYLDN-----QSTYIVVEPILEKENOT-----DKPSILPQFKRN 311
 DB 970 ANGSVDRGAWKYLKRVNNGRYDLNPEVEKRNQTVDTNITTPNDIQADAPS----- 1021

TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1098 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-8

Query Match 5.1%; Score 141; DB 4; Length 1098;
 Best Local Similarity 20.0%; Pred. No. 0.0055;
 Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

```

QY 2 KOLDKKIEKI-----AGIMK--QYGVKRESIVVNKKNAILIYPHGDHHDPI--- 48
DB 203 EDAEVKREELKLFSSKAGLDQEQEHVKKET---SSEENT---QKVDEHYANSLQNL 256
QY 49 -----DEHKPVGIGHSHSNYELFKPEEGVAKKCKGVY--TGEELTNVNVLL 93
DB 257 AOKSLEELDKATTNEQATQVKNQFLENAQKLEIQPLIKETNVKLYKAMSESLQVEKEL 316
QY 94 KNSTFNQNTFLANGQKRVSEFPEPELEKLGINMLVKLITPDGKVLKSVGKVFGEVGV 153
DB 317 KHSEANLEDLVAKSEIVR-----EYEGKL--NQSKNL--PELKOLEEEAHSKLVQVVE 367
QY 154 NI-ANFELDQPLPGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
DB 368 DFRKFKTSEQVTPKRLKRDLAANENNOQKIELTVSPENITVYEGEDVKFTVTA---K 423
QY 198 MASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGKIKLPK 257
DB 424 SDSKT---TLDFSLLTKYNPSVS-----DRISTNKTNT---DNHKAETITKNLK 469
QY 258 LNOGTTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILEKENQTDKPSILPOFKRN 311
DB 470 LNESQTVTLAKDDSGNVKFTF-----TITVQKKEK-----QVPKT 507
QY 312 KAOENSKLDEKV-EPEKTSKVEKEKSLSETGN-----STNSTLEEVPV----- 355
DB 508 PEQKDSKTEKVPQEPKSNKQNLQELIKSAQOQLEKLEKAIKELMEQPEIPSNPEYGIQ 567
QY 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGEV 395
DB 568 KSWESQKEPIQEAITSFKKIIGDSSSKYTYEHYFNKYKSDFMNQLHAQM-----EMLT 622
QY 396 KNAQFTGEAPOGNGENKPSNGKYST-----GTVENQPT-----NK----- 433
DB 623 RKVQYMNKYPDNAEIKKIFESDMKRTKEDNNGVSLNDALKGYEKFYFLTPFNKIQIVD 682
QY 434 PADSLPEAPNEKVPKPNSTNGMLNPSNGVSDPMDLPALEEA-----PAVDPVQ 484
DB 683 DFDKKEVQEQDQAPI-PENSE-----MDQAKKAKIAVSKYMSKVLGDGVH 725
QY 485 EKLERFTASYGLGL-----DSVIFNMD 506
DB 726 QHLQKNKSKIVDLKFLEAIKQQTIFDID 755

```

RESULT 9
 US-08-923-992A-2
 Sequence 2, Application US/08923992A
 Patent No. 6280738
 GENERAL INFORMATION:
 APPLICANT: Tai, Joseph Y.
 APPLICANT: Blake, Milan S.
 TITLE OF INVENTION: No. 6280738-IgA Fc-Binding Forms of the Group B
 TITLE OF INVENTION: Streptococcal Beta Antigens
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.

COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,707
 FILING DATE: 06-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1164 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-2

Query Match 5.1%; Score 141; DB 4; Length 1164;
 Best Local Similarity 20.0%; Pred. No. 0.006;
 Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

```

QY 2 KOLDKKIEKI-----AGIMK--QYGVKRESIVVNKKNAILIYPHGDHHDPI--- 48
DB 240 EDAEVKREELKLFSSKAGLDQEQEHVKKET---SSEENT---QKVDEHYANSLQNL 293
QY 49 -----DEHKPVGIGHSHSNYELFKPEEGVAKKCKGVY--TGEELTNVNVLL 93
DB 294 AOKSLEELDKATTNEQATQVKNQFLENAQKLEIQPLIKETNVKLYKAMSESLQVEKEL 353
QY 94 KNSTFNQNTFLANGQKRVSEFPEPELEKLGINMLVKLITPDGKVLKSVGKVFGEVGV 153
DB 354 KHSEANLEDLVAKSEIVR-----EYEGKL--NQSKNL--PELKOLEEEAHSKLVQVVE 404
QY 154 NI-ANFELDQPLPGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
DB 405 DFRKFKTSEQVTPKRVKRDLAANENNOQKIELTVSPENITVYEGEDVKFTVTA---K 460
QY 198 MASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGKIKLPK 257
DB 461 SDSKT---TLDFSLLTKYNPSVS-----DRISTNKTNT---DNHKAETITKNLK 506
QY 258 LNOGTTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILEKENQTDKPSILPOFKRN 311
DB 507 LNESQTVTLAKDDSGNVKFTF-----TITVQKKEK-----QVPKT 544
QY 312 KAOENSKLDEKV-EPEKTSKVEKEKSLSETGN-----STNSTLEEVPV----- 355
DB 545 PEQKDSKTEKVPQEPKSNKQNLQELIKSAQOQLEKLEKAIKELMEQPEIPSNPEYGIQ 604
QY 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGEV 395
DB 605 KSWESQKEPIQEAITSFKKIIGDSSSKYTYEHYFNKYKSDFMNQLHAQM-----EMLT 559
QY 396 KNAQFTGEAPOGNGENKPSNGKYST-----GTVENQPT-----NK----- 433
DB 660 RKVQYMNKYPDNAEIKKIFESDMKRTKEDNNGVSLNDALKGYEKFYFLTPFNKIQIVD 719
QY 434 PADSLPEAPNEKVPKPNSTNGMLNPSNGVSDPMDLPALEEA-----PAVDPVQ 484
DB 720 DLDDKKEVQEQDQAPI-PENSE-----MDQAKKAKIAVSKYMSKVLGDGVH 762

```


QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 763 OHLOKKNNSKIVDLFKELEAIKQOTIFDID 792

RESULT 10

US-08-923-992A-6

; Sequence 6, Application US/08923992A

; Patent No. 6280738

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Blake, Milan S.

; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B

; TITLE OF INVENTION: Streptococcal Beta Antigens

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, N.W., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/923,992A

; FILING DATE: 05-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/024,707

; FILING DATE: 06-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1128 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-923-992A-6

Query Match 5.1%; Score 140; DB 4; Length 1128;

Best Local Similarity 19.8%; Pred. No. 0.007;

Matches 125; Conservative 96; Mismatches 207; Indels 202; Gaps 32;

QY 2 KLDLKIEEKI-----AGIMKQY--GVKRESIVVNKKNAILIYPHGDHHDPI---48

Db 204 EDAEVKVRLEGLKFSSTRAGLQDEIHEHVKKET---SSEENT---QKVDEHYANSLQNL 257

QY 49 -----DEHKVPVIGHSNHYELFKPEGVAKKEGKVI--TGEELTNVNNLL 93

Db 258 AOKSLEELDKATTNEQATQYKNOFLNNAQKLKEMQPLIKETNVKLYKAMSESLEQVEKEL 317

QY 94 KNSTFNQNTTANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKVSQKVFGEVGV 153

Db 318 KHNSANLEDLVAKSKEIVR-----EYEGKL--NQSKNL--PELKQLEEEAHSKLQOVVE 368

QY 154 NI-ANFELDQYPLFGQTFKYTIASKD-----YPE-VSYDG---TFTVPTSLAYK 197

Db 369 DFRKKFKTSEQVTPKRVKRDLAANNQKIELTVSPENITVIEGEDVKFTVTA---K 424

QY 198 MASQTIYPPHAGDTYLRVNPQAVPKGTDALVRFVDEHGNAYLENNYKVGCIKPIPK 257

Db 425 SDSKNT--TLDFSDLLTKYNPVS-----DRISTNRYKNT--DNRKIADITIKNLK 470

QY 258 LNQGTIRT-----AGNKIPVTTFMANAYLDNQSTVIVVEPILEKENQTDKPSILPQFARN 311
Db 471 LNESQTVTLKARDDSGNVYKTF-----TITVQKKEK-----QVPEKT 508
QY 312 KAQENSKLDEKV-EPKPTSEKVEKEKLSETGN-----STNSTLEEYPTV----- 355
Db 509 PEQKDSKTEEKVPQEPKSNQNDKNQLOELIKSAQOELEKLEKAIKELMEQPEIPSPNPEYCIQ 568
QY 356 -----DPVOEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGEVIR 395
Db 569 KSIMESQKEPIQEAITSFKKIIGDSSSKYVTEHYFNPKYKSDFMNYQLHAQM-----EMLT 623
QY 396 KNAADFTGEAPQNGENKPSENGKYST-----GTVENQPT-----NK----- 433
Db 624 RNVGYMMKYPDNAEIKKIFESDMARKEDNNGVSGLENDALKGYPEKYLTFPFNKIKQIVD 683
QY 434 PADSLPEAPNEKPVKPNSTNDGMLNPEGNVSGSDPMLDPALEEA-----PAVDVQV 484
Db 684 DLDDKVEQDPAPI-PENSE-----MDQAKEKAKIAVSKYMSKVLDOGVH 726
QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 727 OHLOKKNNSKIVDLFKELEAIKQOTIFDID 756

RESULT 11

US-08-923-992A-4

; Sequence 4, Application US/089233992A

; Patent No. 6280738

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Blake, Milan S.

; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B

; TITLE OF INVENTION: Streptococcal Beta Antigens

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, N.W., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/923,992A

; FILING DATE: 05-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/024,707

; FILING DATE: 06-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1104 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-923-992A-4

Query Match

5.1%; Score 139; DB 4; Length 1104;

Best Local Similarity 19.8%; Pred. No. 0.0082;

Matches 125; Conservative 95; Mismatches 208; Indels 202; Gaps 32;


```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-714-481-2

Query Match          5.0%; Score 138.5; DB 1; Length 984;
Best Local Similarity 17.0%; Pred. No. 0.0076;
Matches 101; Conservative 95; Mismatches 176; Indels 221; Gaps 25;

QY 1 MKDLDDKIEKIAGIMQGVKRESIVVNEKKNALIIYPHGD-HHHADPIDEHKPVGI--- 56
Db 154 LSNIDKELNHQKSOVEK---MAEQGIINEDKDSMLKIEDIRKQAQADKEDAEVRVQ 210
QY 57 --GHSNVELFKPEGVAKKGNKYVGEELTNVNLKNSTFNQNTLANGOKRVSF 114
Db 211 LEEEAHSLK-----KQVDEFRKKFKTSEQVTPKKRVKRDLAANE-----NNQKI 258
QY 115 SPPELEKKGKGINMLVLIPTDGVLEKVGKVGEGVGNIANEELDQPLPGQTFKYTI 174
Db 259 TVSPE-----NITVE-----FSDLLTKYNPSVS-----DRIST 306
QY 175 ASKDYPEVSDGTFTVPTSLAYKMASOTIYFFHAGDTYLRVNPQFVAPKGTDLVRVFD 234
Db 278 TAKSDSKTTLD-----FSDLLTKYNPSVS-----DRIST 306
QY 235 EFHGNAYLENNYKVGKIKPIPKLNQGTTRT-----AGNKIPVTFMANAYLDNQSTIY 288
Db 307 NYKNT---DNHKAETIKNLKINESQTVTKAKDDSGNVVEKTF----- 349
QY 289 EYPILEKENQTDKPSILPOFKENKAQENSKLDEKV-EPEKTSKVEKEKLSGTGN----- 342
Db 350 TITVQKKEK-----QVPKTPQKDSKTEEVKVPQPKSNDKNQLOELIKSAQOELEK 401
QY 343 -STSNTLEEVTV-----DPQVEKVAKF-----AESYGNK 372
Db 402 LEKALKEKMEQPEIPSNPEYGIQKSIWESQKPEIQEATISFKKIIGDSSSKYYTEHYENK 461
QY 373 LENVLNFMGDTIELYLPSPGEVKKKNMADFTGAPQNGENKPSNGKYST-----GTVEN 427
Db 462 YKSDFNWQLHAQM-----EMLTRKVVQYMKYKYPDPAEIKKIFESDMKRTKEDNNGVSGLEN 516
QY 428 QPTE-----NK-----PADSLPEAPNEKPKVPKPNSTNDNGMLNPEGNGSDPML 470
Db 517 DALKGYFEKRYFLTPFNKIKQIVDDLDKKVQEQDQAPPI-PENSE-----M 559
QY 471 DPALKEA-----PAVDPPQKLEKFTASYGLGL-----DSVIFNMD 506
Db 560 DOAKERAKTAVSKYMSKVLGDGVHQHLQKNNSQIVDLFKELEAIKQQTIFDID 612

RESULT 14

```

```

PCT-US95-061111-2
; Sequence 2, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-061111-2

```

```

Query Match          5.0%; Score 138.5; DB 5; Length 984;
Best Local Similarity 17.0%; Pred. No. 0.0076;
Matches 101; Conservative 95; Mismatches 176; Indels 221; Gaps 25;

QY 1 MKDLDDKIEKIAGIMQGVKRESIVVNEKKNALIIYPHGD-HHHADPIDEHKPVGI--- 56
Db 154 LSNIDKELNHQKSOVEK---MAEQGIINEDKDSMLKIEDIRKQAQADKEDAEVRVQ 210
QY 57 --GHSNVELFKPEGVAKKGNKYVGEELTNVNLKNSTFNQNTLANGOKRVSF 114
Db 211 LEEEAHSLK-----KQVDEFRKKFKTSEQVTPKKRVKRDLAANE-----NNQKI 258
QY 115 SPPELEKKGKGINMLVLIPTDGVLEKVGKVGEGVGNIANEELDQPLPGQTFKYTI 174
Db 259 TVSPE-----NITVE-----GEDVRKTV 277
QY 175 ASKDYPEVSDGTFTVPTSLAYKMASOTIYFFHAGDTYLRVNPQFVAPKGTDLVRVFD 234
Db 278 TAKSDSKTTLD-----FSDLLTKYNPSVS-----DRIST 306
QY 235 EFHGNAYLENNYKVGKIKPIPKLNQGTTRT-----AGNKIPVTFMANAYLDNQSTIY 288
Db 307 NYKNT---DNHKAETIKNLKINESQTVTKAKDDSGNVVEKTF----- 349

```

```
QY 289 EPTILEKENTQDKSIIPQKRNKAQENSKLDEKV-EERKTSKVEKEKLESETGN----- 342
DB 350 TITVQKREK-----QVFKTPEQKDSKTEKVPQEPKSNKQNLQELIKSAQOELEK 401
QY 343 -STNSSTLEEVPTV-----DPVQEKVAKF-----AESYGMK 372
DB 402 LEKAELMEQPEPSPNEXGKIOSWESOKEPQIAITSFKKLIQDSSSKYTYEHFNK 461
QY 373 LENVLNMDGTIELYLPSEGVKIKNMAFTGEAPQNGENKPKSENGKYST-----GTVEN 427
DB 462 YKSDFMNYQLHAQM-----EMLTRKVYQVYNNKYPDNAETKKIFESDMKRTKEDNYGSLFN 516
QY 428 QPTE-----NK-----PADSLPEAPNEKPKVPKPNSTDNMNLNPEGNVGSDPML 470
DB 517 DALGYPEKFTLTFNKAIKOIVDDLDKKVEQDQAPFI-PENSE-----M 559
QY 471 DPALAEA-----PAVDPVQOEKLEKFTASYGLGL-----DSVIFNMD 506
DB 560 DOAKEKAKIAVSKYMSKVLGDGVHQHLQKNNKXIVDLFKLEAIKOOTIFDID 612

RESULT 15
US-08-589-756-1
: Sequence 1, Application US/08589756
: Patent No. 5846547
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,756
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-589-756-1

Query Match 4.88; Score 132.5; DB 2; Length 1164;
Best Local Similarity 23.48; Pred.No. 0.03;
Matches 91; Conservative 43; Mismatches 144; Indels 111; Gaps 19;

QY 17 KOYGVKRESIVYNNKNAIIPRGDHHHADPIDEHKPVGIGHSHSNYELFKPEGVAKKE 76
DB 805 RHYVIHRHA---NGKPYAAISPNGDGNR-DYVQFH-----GTELRNAKNLVAE--VLDKE 853
QY 77 GNKYVTGTELTNNVNLKN---STFNQNF-----TLANGOKRVSFSPFP-- 118
DB 854 GNVVMTSEVTEQVYNNVNDLASTLGTREISRWGDKDKAKVWANGTYTYRVRTPIIS 913
QY 119 --ELEKLGINMLVKLITPBGVLEKYSKVGFGVGNIANF-----ELDQP 163
DB 914 SGAKEQHTDQDVIVDNTPE-----VATSATFSTEDRRUTLASKPQTSOP 958
QY 164 YLPQGTFKYITASKDYPEVSY-----DGTITVP-----TSLAYKMASQIIFYPPFHAG 210
DB 959 -VYRERAIYTMDELDLETTETYSFNGEDGTFLEPEATFEGATVPLKMSDFTYVVEDMAG 1017
QY 211 D-TYLRVNPQFAVPKGDALVRVDFEPHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNK 269
DB 1018 NITVTPV-----TKLL-----EGHSNKPQDGDQDQAPDKKPKETKPEQDGSQAPDK 1063
QY 270 IPVTFMANAYLDNSTIVIEVPILEKENQI--DKPSILPQFKENKAQENSKLDEKVEEPK 327
DB 1064 KPET-----KPGQDGGSGQTPDKKPKETKPE-KDSSGQTPGKTFQKGQPSR 1106
```

```
QY 328 TSEKVEKEKLESETGNSTSTLEEVPTVD 356
DB 1107 TLEK-----RSKRALATKASTROOLPTTN 1131

Search completed: May 13, 2003, 13:57:52
Job time : 15.1843 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 13, 2003, 13:53:57 ; Search time 13,0507 Seconds
(without alignments)
3723.142 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKIEKIGIMKQYGVKRESIVVNNKKNALITYPHGDHHDHPIDHKPVGTGHS

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues
Total number of hits satisfying chosen parameters: 349150
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2746	100.0	569	9	US-09-884-465A-235 Sequence 235, App
2	2746	100.0	569	9	US-09-884-465A-255 Sequence 255, App
3	2746	100.0	627	9	US-09-884-465A-366 Sequence 366, App
4	2746	100.0	633	9	US-09-884-465A-350 Sequence 350, App
5	2746	100.0	633	9	US-09-884-465A-351 Sequence 351, App
6	2746	100.0	633	9	US-09-884-465A-362 Sequence 362, App
7	2746	100.0	894	9	US-09-884-465A-337 Sequence 337, App
8	2746	100.0	895	9	US-09-884-465A-347 Sequence 347, App
9	2746	100.0	900	9	US-09-884-465A-335 Sequence 335, App
10	2746	100.0	900	9	US-09-884-465A-339 Sequence 339, App
11	2746	100.0	900	9	US-09-884-465A-341 Sequence 341, App
12	2746	100.0	901	9	US-09-884-465A-343 Sequence 343, App
13	2746	100.0	901	9	US-09-884-465A-345 Sequence 345, App
14	2746	100.0	913	9	US-09-884-465A-384 Sequence 384, App
15	2746	100.0	999	9	US-09-884-465A-376 Sequence 376, App
16	2746	100.0	999	9	US-09-884-465A-377 Sequence 377, App
17	2746	100.0	1039	9	US-09-884-465A-6 Sequence 6, Appl
18	2746	100.0	1039	9	US-09-884-465A-6 Sequence 383, App
19	2746	100.0	1126	9	US-09-884-465A-383 Sequence 383, App

ALIGNMENTS

RESULT 1
US-09-884-465A-235
Sequence 235, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 235
LENGTH: 569
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-235

Query Match 100.0%; Score 2746; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.3e-179;
Matches 528; Conservative 0; Mismatches 0; Gaps 0;

Qy	1	MKDLKKIEKIGIMKQYGVKRESIVVNNKKNALITYPHGDHHDHPIDHKPVGTGHS	60
Db	42	MKDLKKIEKIGIMKQYGVKRESIVVNNKKNALITYPHGDHHDHPIDHKPVGTGHS	101
Qy	61	SNYELFKPEEGVAKKGNKVTGTELTNNVNNLKNSTFNNQNTLANGOKRVSFSPPEL	120
Db	102	SNYELFKPEEGVAKKGNKVTGTELTNNVNNLKNSTFNNQNTLANGOKRVSFSPPEL	161
Qy	121	EKXLGINLWKLITPDGKLVLEKVGKGVGCVGNINAFELDQPLPGQTFKYTIASKDYP	180
Db	162	EKXLGINLWKLITPDGKLVLEKVGKGVGCVGNINAFELDQPLPGQTFKYTIASKDYP	221

20	2746	100.0	1139	9	US-09-884-465A-380	Sequence 380, App
21	2746	100.0	1152	9	US-09-884-465A-379	Sequence 379, App
22	2746	100.0	1238	9	US-09-884-465A-381	Sequence 381, App
23	2746	100.0	1365	9	US-09-884-465A-382	Sequence 382, App
24	2746	100.0	1378	9	US-09-884-465A-378	Sequence 378, App
25	2737	99.7	633	9	US-09-884-465A-352	Sequence 352, App
26	2737	99.7	633	9	US-09-884-465A-354	Sequence 354, App
27	2737	99.7	633	9	US-09-884-465A-377	Sequence 357, App
28	2737	99.7	633	9	US-09-884-465A-358	Sequence 358, App
29	2737	99.7	633	9	US-09-884-465A-359	Sequence 359, App
30	2737	99.7	633	9	US-09-884-465A-360	Sequence 360, App
31	2737	99.7	900	9	US-09-884-465A-333	Sequence 333, App
32	2737	99.7	900	9	US-09-884-465A-334	Sequence 334, App
33	2737	99.7	906	9	US-09-884-465A-332	Sequence 332, App
34	2737	99.7	906	9	US-09-884-465A-369	Sequence 369, App
35	2737	99.7	906	9	US-09-884-465A-370	Sequence 370, App
36	2737	99.7	906	9	US-09-884-465A-373	Sequence 373, App
37	2736	99.6	569	9	US-09-884-465A-356	Sequence 356, App
38	2736	99.6	633	9	US-09-884-465A-363	Sequence 363, App
39	2736	99.6	633	9	US-09-884-465A-364	Sequence 364, App
40	2727	99.3	633	9	US-09-884-465A-353	Sequence 353, App
41	2727	99.3	633	9	US-09-884-465A-355	Sequence 355, App
42	2727	99.3	633	9	US-09-884-465A-365	Sequence 365, App
43	2727	99.3	906	9	US-09-884-465A-371	Sequence 371, App
44	2727	99.3	906	9	US-09-884-465A-372	Sequence 372, App
45	2714	98.8	840	9	US-09-884-465A-10	Sequence 10, Appl

181	QY	EVSYDGTFTVPTSLAYKKMASQTIFFPFHAGDTYLRVNQFOFAPVPGKTDALVRVDFDEFGHNA	240
182	QY		241
183	QY		242
184	QY		243
185	QY		244
186	QY		245
187	QY		246
188	QY		247
189	QY		248
190	QY		249
191	QY		250
192	QY		251
193	QY		252
194	QY		253
195	QY		254
196	QY		255
197	QY		256
198	QY		257
199	QY		258
200	QY		259
201	QY		260
202	QY		261
203	QY		262
204	QY		263
205	QY		264
206	QY		265
207	QY		266
208	QY		267
209	QY		268
210	QY		269
211	QY		270
212	QY		271
213	QY		272
214	QY		273
215	QY		274
216	QY		275
217	QY		276
218	QY		277
219	QY		278
220	QY		279
221	QY		280
222	Db	EVSYDGTFTVPTSLAYKKMASQTIFFPFHAGDTYLRVNQFOFAPVPGKTDALVRVDFDEFGHNA	281
223	Db		282
224	Db		283
225	Db		284
226	Db		285
227	Db		286
228	Db		287
229	Db		288
230	Db		289
231	Db		290
232	Db		291
233	Db		292
234	Db		293
235	Db		294
236	Db		295
237	Db		296
238	Db		297
239	Db		298
240	Db		299
241	QY	YLENNYKVGEIKLPKLNQGTTRTAGNKIPVTEMANAYLDNQSTIYIVEVPILEKENQTD	300
242	QY		301
243	QY		302
244	QY		303
245	QY		304
246	QY		305
247	QY		306
248	QY		307
249	QY		308
250	QY		309
251	QY		310
252	QY		311
253	QY		312
254	QY		313
255	QY		314
256	QY		315
257	QY		316
258	QY		317
259	QY		318
260	QY		319
261	QY		320
262	QY		321
263	QY		322
264	QY		323
265	QY		324
266	QY		325
267	QY		326
268	QY		327
269	QY		328
270	QY		329
271	QY		330
272	QY		331
273	QY		332
274	QY		333
275	QY		334
276	QY		335
277	QY		336
278	QY		337
279	QY		338
280	QY		339
281	QY		340
282	Db	YLENNYKVGEIKLPKLNQGTTRTAGNKIPVTEMANAYLDNQSTIYIVEVPILEKENQTD	341
283	Db		342
284	Db		343
285	Db		344
286	Db		345
287	Db		346
288	Db		347
289	Db		348
290	Db		349
291	Db		350
292	Db		351
293	Db		352
294	Db		353
295	Db		354
296	Db		355
297	Db		356
298	Db	</	

```

RESULT 3
US-09-884-465A-366
; Sequence 366, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 366
; LENGTH: 627
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; US-09-884-465A-366

```

	Query Match	100.0%	Score 2746;	DB 9;	Length 569;
	Best Local Similarity	100.0%;	Pred. No. 1.3e-178;		
	Matches 528; Conservative	0;	Mismatches	0;	Gaps 0;
Qy	1	MKDLDKKTEEKTAGTMMKQYGVKRSTVYNKEKNAIYPHGDDHHADPIDEHKPVGIGHSH	60		
Dd	42	MKDLDKKLEETAGTMMKQYGVKRSTVYNKEKNAIYPHGDDHHADPIDEHKPVGIGHSH	101		
Qy	61	SNVELFKPEEGVAKREGNKVTGELTNVNVLKLNSTFNQNFTLANGOKRVSFSEFPPEL	120		
Dd	102	SNVELFKPEEGVAKREGNKVTGELTNVNVLKLNSTFNQNFTLANGOKRVSFSEFPPEL	161		
Qy	121	EKKLGINMLVKILITPDGKVLEKSGKVFGEVGNIANFELDQQPLPGQTFKYTIASKDYP	180		
Dd	162	EKKLGINMLVKILITPDGKVLEKSGKVFGEVGNIANFELDQQPLPGQTFKYTIASKDYP	221		
Qy	181	EVSVDGTFTVPNTSLAYKWASQTIFYPPHAGDTYLVRNPQFAVPGKTIDALVRVDEFHGNA	240		
b	222	EVSVDGTFTVPNTSLAYKWASQTIFYPPHAGDTYLVRNPQFAVPGKTIDALVRVDEFHGNA	281		

us-09-471-255-10.rapb

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 528
Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 633

RESULT 6

US-09-884-465A-361
; Sequence 361, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 361
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-361

Query Match 100.0%; Score 2746; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.4e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
Db 106 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 165
QY 61 SNYELFKPEEGVAKKGNKVYTGELTNVYNLLKNSFTNNQNTFLANGQKRVSFSPPEL 120
Db 166 SNYELFKPEEGVAKKGNKVYTGELTNVYNLLKNSFTNNQNTFLANGQKRVSFSPPEL 225
QY 121 EKKLGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGTFFKYTIASKDYP 180
Db 226 EKKLGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGTFFKYTIASKDYP 285
QY 181 EVSDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 240
Db 286 EVSDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 345
QY 241 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 300
Db 346 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 405
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLESTGNSTNSLTLEEVPTDPVQE 360
Db 406 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLESTGNSTNSLTLEEVPTDPVQE 465
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKNAADFTGEAPQNGENKPSGENKV 420
Db 466 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKNAADFTGEAPQNGENKPSGENKV 525
QY 421 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
Db 526 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEAPAV 585
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 528
Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 633

RESULT 8
US-09-884-465A-337
; Sequence 337, Application US/09884465A
; Publication No. US20030077293A1

RESULT 7
US-09-884-465A-362
; Sequence 362, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 362
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-362

Query Match 100.0%; Score 2746; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.4e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
Db 106 MKDLKKIEEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGHSH 165
QY 61 SNYELFKPEEGVAKKGNKVYTGELTNVYNLLKNSFTNNQNTFLANGQKRVSFSPPEL 120
Db 166 SNYELFKPEEGVAKKGNKVYTGELTNVYNLLKNSFTNNQNTFLANGQKRVSFSPPEL 225
QY 121 EKKLGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGTFFKYTIASKDYP 180
Db 226 EKKLGINMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGTFFKYTIASKDYP 285
QY 181 EVSDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 240
Db 286 EVSDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 345
QY 241 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 300
Db 346 YLENNYKVGGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 405
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLESTGNSTNSLTLEEVPTDPVQE 360
Db 406 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLESTGNSTNSLTLEEVPTDPVQE 465
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKNAADFTGEAPQNGENKPSGENKV 420
Db 466 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKNAADFTGEAPQNGENKPSGENKV 525
QY 421 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
Db 526 STGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEAPAV 585
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 528
Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFTA 633

RESULT 8
US-09-884-465A-337
; Sequence 337, Application US/09884465A
; Publication No. US20030077293A1

; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 337
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-337

Query Match 100.0%; Score 2746; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNEKKNATIIYPHGDHHDADPIDEHKPVGIGHSH 60
Db 367 MKDLKKIEEKIAGIMKQYGVKRESIVVNEKKNATIIYPHGDHHDADPIDEHKPVGIGHSH 426
QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVYLLKNSTFNQNTFLANGQKRVFSFPPEL 120
Db 427 SNYELFKPEEGVAKKEGKNVYTGELTNVYLLKNSTFNQNTFLANGQKRVFSFPPEL 486
QY 121 EKKGINMLVKLITPDGKYLEKVGSKVFGGCVGNIANFELDDQYLPQGTFKYTIASKDYP 180
Db 487 EKKGINMLVKLITPDGKYLEKVGSKVFGGCVGNIANFELDDQYLPQGTFKYTIASKDYP 546
QY 181 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 547 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 606
QY 241 YLENNYKVGEEKIPLKNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPVILEKENQTD 300
Db 607 YLENNYKVGEEKIPLKNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPVILEKENQTD 666
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPVDPVQE 360
Db 667 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPVDPVQE 726
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQNGENKPSENGKV 420
Db 727 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQNGENKPSENGKV 786
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 480
Db 787 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 846
QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVYIKKNLSDFIA 528
Db 847 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVYIKKNLSDFIA 894

RESULT 9
US-09-884-465A-347
; Sequence 347, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis

; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 347
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-347

Query Match 100.0%; Score 2746; DB 9; Length 895;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNEKKNATIIYPHGDHHDADPIDEHKPVGIGHSH 60
Db 368 MKDLKKIEEKIAGIMKQYGVKRESIVVNEKKNATIIYPHGDHHDADPIDEHKPVGIGHSH 427
QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVYLLKNSTFNQNTFLANGQKRVFSFPPEL 120
Db 428 SNYELFKPEEGVAKKEGKNVYTGELTNVYLLKNSTFNQNTFLANGQKRVFSFPPEL 487
QY 121 EKKGINMLVKLITPDGKYLEKVGSKVFGGCVGNIANFELDDQYLPQGTFKYTIASKDYP 180
Db 488 EKKGINMLVKLITPDGKYLEKVGSKVFGGCVGNIANFELDDQYLPQGTFKYTIASKDYP 547
QY 181 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 548 EVSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 607
QY 241 YLENNYKVGEEKIPLKNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPVILEKENQTD 300
Db 608 YLENNYKVGEEKIPLKNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPVILEKENQTD 667
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPVDPVQE 360
Db 668 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPVDPVQE 727
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQNGENKPSENGKV 420
Db 728 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQNGENKPSENGKV 787
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 480
Db 788 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 847
QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVYIKKNLSDFIA 528
Db 848 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVYIKKNLSDFIA 895

RESULT 10
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A

```
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 335
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-335

Query Match          100.0%; Score 2746; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDDHHADPIDEHKPVGIGHSH 60
Db 373 MKDLKKIEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDDHHADPIDEHKPVGIGHSH 432
QY 61 SNYELFKPEGVAKKGNKYVTGEEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 120
Db 433 SNYELFKPEGVAKKGNKYVTGEEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 492
QY 121 EKKLGINMLVKLITPDGKLVKSVGEGVGNANFELDQPYLPQOTFKYTIASKDYP 180
Db 493 EKKLGINMLVKLITPDGKLVKSVGEGVGNANFELDQPYLPQOTFKYTIASKDYP 552
QY 181 EYSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
Db 553 EYSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 612
QY 241 YLENNYKVGKIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILKEKNTD 300
Db 613 YLENNYKVGKIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILKEKNTD 672
QY 301 KPSILPQFKRNAQENSKLDEKVEEPEKTSKVEKELSETGNSNSTLEEVPITVDPVQE 360
Db 673 KPSILPQFKRNAQENSKLDEKVEEPEKTSKVEKELSETGNSNSTLEEVPITVDPVQE 732
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSENGKV 420
Db 733 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSENGKV 792
QY 421 STGTVENOPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEEAPAV 480
Db 793 STGTVENOPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEEAPAV 852
QY 481 DPVQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 528
Db 853 DPVQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 900

RESULT 11
US-09-884-465A-339
; Sequence 339, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1

Query Match          100.0%; Score 2746; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDDHHADPIDEHKPVGIGHSH 60
Db 373 MKDLKKIEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDDHHADPIDEHKPVGIGHSH 432
QY 61 SNYELFKPEGVAKKGNKYVTGEEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 120
Db 433 SNYELFKPEGVAKKGNKYVTGEEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 492
QY 121 EKKLGINMLVKLITPDGKLVKSVGEGVGNANFELDQPYLPQOTFKYTIASKDYP 180
Db 493 EKKLGINMLVKLITPDGKLVKSVGEGVGNANFELDQPYLPQOTFKYTIASKDYP 552
QY 181 EYSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
Db 553 EYSVDGTFVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 612
QY 241 YLENNYKVGKIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILKEKNTD 300
Db 613 YLENNYKVGKIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILKEKNTD 672
QY 301 KPSILPQFKRNAQENSKLDEKVEEPEKTSKVEKELSETGNSNSTLEEVPITVDPVQE 360
Db 673 KPSILPQFKRNAQENSKLDEKVEEPEKTSKVEKELSETGNSNSTLEEVPITVDPVQE 732
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSENGKV 420
Db 733 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSENGKV 792
QY 421 STGTVENOPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEEAPAV 480
Db 793 STGTVENOPTENKPADSLPEAPNEKPKPENSTDNGLNPEGNVGSDDPMLDPALEEAPAV 852
QY 481 DPVQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 528
Db 853 DPVQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 900

RESULT 12
US-09-884-465A-341
; Sequence 341, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 341
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

Query Match 100.0%; Score 2746; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
DB 374 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 433

QY 61 SNYELFPEEGVAKKEGKNTVTELTNVNLLKNSTFNNQNTFLANGOKRVSFSPPEL 120
DB 434 SNYELFPEEGVAKKEGKNTVTELTNVNLLKNSTFNNQNTFLANGOKRVSFSPPEL 493

QY 121 EKKLGIMLVKLIITPDGKVLKESKGVFEGVGNIANFELDPQYLPQGTFFKTYIASKDYP 180
DB 494 EKKLGIMLVKLIITPDGKVLKESKGVFEGVGNIANFELDPQYLPQGTFFKTYIASKDYP 553

QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
DB 554 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 613

QY 241 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEVPILKENQTD 300
DB 614 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEVPILKENQTD 673

QY 301 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEYVTPVDPVOE 360
DB 674 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEYVTPVDPVOE 733

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENGKV 420
DB 734 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENGKV 793

QY 421 STGTVENOPTENKPADSLPEAPNEKPKVPENSTDNGLMNPENGVGSDPMLDPALEAPAV 480
DB 794 STGTVENOPTENKPADSLPEAPNEKPKVPENSTDNGLMNPENGVGSDPMLDPALEAPAV 853

QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIKKNLSDFIA 528
DB 854 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIKKNLSDFIA 901

RESULT 13
US-09-884-465A-343
Sequence 343, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
CURRENT FILING DATE: 2001-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 343
LENGTH: 901
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

Query Match 100.0%; Score 2746; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

Query Match 100.0%; Score 2746; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
DB 374 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 433

QY 61 SNYELFPEEGVAKKEGKNTVTELTNVNLLKNSTFNNQNTFLANGOKRVSFSPPEL 120
DB 434 SNYELFPEEGVAKKEGKNTVTELTNVNLLKNSTFNNQNTFLANGOKRVSFSPPEL 493

QY 121 EKKLGIMLVKLIITPDGKVLKESKGVFEGVGNIANFELDPQYLPQGTFFKTYIASKDYP 180
DB 494 EKKLGIMLVKLIITPDGKVLKESKGVFEGVGNIANFELDPQYLPQGTFFKTYIASKDYP 553

QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
DB 554 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 613

QY 241 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEVPILKENQTD 300
DB 614 YLENNYKVGEEKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOSTYIIVEVPILKENQTD 673

QY 301 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEYVTPVDPVOE 360
DB 674 KPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEYVTPVDPVOE 733

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENGKV 420
DB 734 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENGKV 793

QY 421 STGTVENOPTENKPADSLPEAPNEKPKVPENSTDNGLMNPENGVGSDPMLDPALEAPAV 480
DB 794 STGTVENOPTENKPADSLPEAPNEKPKVPENSTDNGLMNPENGVGSDPMLDPALEAPAV 853

QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIKKNLSDFIA 528
DB 854 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIKKNLSDFIA 901

RESULT 14
US-09-884-465A-345
Sequence 345, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
CURRENT FILING DATE: 2001-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 345
LENGTH: 901
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

Query Match 100.0%; Score 2746; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 60
DB 374 MKDLKKIEEKIAGIMKQYGVKRESIVVWNEKNAIYPHGDHHDADPIDEHKPVGIGHSH 433

QY 61 SNYELFKPEGVAKKREGNKVYTGELTNVNVLLKKNSTNNQNTLANGOKRVVSFPPEL 120
DB 434 SNYELFKPEGVAKKREGNKVYTGELTNVNVLLKKNSTNNQNTLANGOKRVVSFPPEL 493
QY 121 EKKGLINMLVKLITPDGKLVLEKVSQVFGGNGNTANFELDQPYLPQGTFFYTTASKDYP 180
DB 494 EKKGLINMLVKLITPDGKLVLEKVSQVFGGNGNTANFELDQPYLPQGTFFYTTASKDYP 553
QY 181 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
DB 554 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 613
QY 241 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 300
DB 614 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 673
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTNSLLEEVPTVDPVOE 360
DB 674 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTNSLLEEVPTVDPVOE 733
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 420
DB 734 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 793
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPNSENGVSDPMLDPALEAPAV 480
DB 794 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPNSENGVSDPMLDPALEAPAV 853
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 528
DB 854 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 901

RESULT 15

US-09-884-465A-384
; Sequence 384, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 384
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)-(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (344)..(344)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-384

Query Match

Best Local Similarity 100.0%; Score 2746; DB 9; Length 913;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQXGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSH 60
DB 386 MKDLKKIEEKIAGIMKQXGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSH 445
QY 61 SNYELFKPEGVAKKREGNKVYTGELTNVNVLLKKNSTNNQNTLANGOKRVVSFPPEL 120
DB 446 SNYELFKPEGVAKKREGNKVYTGELTNVNVLLKKNSTNNQNTLANGOKRVVSFPPEL 505
QY 121 EKKGLINMLVKLITPDGKLVLEKVSQVFGGNGNTANFELDQPYLPQGTFFYTTASKDYP 180
DB 506 EKKGLINMLVKLITPDGKLVLEKVSQVFGGNGNTANFELDQPYLPQGTFFYTTASKDYP 565
QY 181 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
DB 566 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 625
QY 241 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 300
DB 626 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 685
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTNSLLEEVPTVDPVOE 360
DB 686 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTNSLLEEVPTVDPVOE 745
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 420
DB 746 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 805
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPNSENGVSDPMLDPALEAPAV 480
DB 806 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPNSENGVSDPMLDPALEAPAV 865
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 528
DB 866 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 913

Search completed: May 13, 2003, 13:58:53

Job time : 15.0507 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:49:42 ; Search time 35.8944 Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAYALNQHRSOENKNNRVS.....GTIELRLPSGEVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4396	100.0	840	21	AA12753
2	4396	100.0	840	23	AAU76151
3	4276.5	97.3	1019	21	AA12722
4	4276.5	97.3	1019	21	AA12751
5	4276.5	97.3	1019	23	AAU84021
6	4276.5	97.3	1039	23	AA12715
7	4276.5	97.3	1039	23	AAU75932
8	4274.5	97.2	1019	21	AA12750
9	4268.5	97.1	1019	21	AA12748
10	4268.5	97.1	1019	21	AA12752

11	4268.5	97.1	1019	21	AA12753	Streptococcus pneu
12	4264.5	97.0	1019	21	AA12749	Streptococcus pneu
13	3402.5	77.4	807	23	AAU83997	Truncated variant
14	3402.5	77.4	1152	23	AAU84054	S. pneumoniae deri
15	3402.5	77.4	1238	23	AAU84056	S. pneumoniae deri
16	3402.5	77.4	1365	23	AAU84057	S. pneumoniae deri
17	3402.5	77.4	1378	23	AAU84053	S. pneumoniae deri
18	3241	73.7	644	23	AAU83996	Truncated variant
19	3237	73.6	907	23	AAU84051	S. pneumoniae deri
20	3228	73.4	907	23	AAU84053	S. pneumoniae deri
21	3219	73.2	907	23	AAU84053	S. pneumoniae deri
22	3215	73.1	901	23	AAU84067	S. pneumoniae deri
23	3202.5	72.9	632	23	AAU83998	Truncated variant
24	3200.5	72.8	902	23	AAU84076	S. pneumoniae deri
25	3193.5	72.6	632	23	AAU84002	Truncated variant
26	3193.5	72.6	632	23	AAU84006	Truncated variant
27	3193.5	72.6	632	23	AAU84013	Truncated variant
28	3193.5	72.6	632	23	AAU84014	Truncated variant
29	3191	72.6	901	23	AAU84088	S. pneumoniae deri
30	3184.5	72.4	632	23	AAU84003	Truncated variant
31	3184.5	72.4	632	23	AAU84004	Truncated variant
32	3184.5	72.4	632	23	AAU84009	Truncated variant
33	3184.5	72.4	632	23	AAU84012	Truncated variant
34	3184.5	72.4	901	23	AAU84066	S. pneumoniae deri
35	3184.5	72.4	907	23	AAU84080	S. pneumoniae deri
36	3184.5	72.4	907	23	AAU84084	S. pneumoniae deri
37	3183.5	72.4	632	23	AAU84015	Truncated variant
38	3183.5	72.4	632	23	AAU84016	Truncated variant
39	3182.5	72.4	902	23	AAU84078	S. pneumoniae deri
40	3182	72.4	901	23	AAU84072	S. pneumoniae deri
41	3182	72.4	901	23	AAU84074	S. pneumoniae deri
42	3178.5	72.3	895	23	AAU84080	S. pneumoniae deri
43	3178	72.3	895	23	AAU84070	S. pneumoniae deri
44	3174.5	72.2	632	23	AAU84007	Truncated variant
45	3174.5	72.2	632	23	AAU84017	Truncated variant

ALIGNMENTS

RESULT 1
AA12721
ID AA12721 standard; Protein: 840 AA.
XX AA12721;
AC AA12721;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
OS
XX
XX PN WO200039299-A2.
XX
XX PD 06-JUL-2000.
XX
XX PF 20-DEC-1999; 99WO-CA01218.
XX
XX PR 23-DEC-1998; 98US-0113800.
XX
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX DR N-PSDB; AAA65738.
XX
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -

```
XX PS Claim 18; Fig 19; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae Sp63 BVH-3 protein antigen.
XX SQ Sequence 840 AA;

Query Match 100.0%; Score 4396; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 1.le-261;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKREGIOAEQIVIKITDQGYV 60
DB 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKREGIOAEQIVIKITDQGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKADIYNEVKGVIIVKDGKYYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKADIYNEVKGVIIVKDGKYYVYLKD 120
QY 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVFNPADIEDTGA 180
DB 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVFNPADIEDTGA 180
QY 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGNNQPSQSYSTPSPSLPINPGTSHEKH 240
DB 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGNNQPSQSYSTPSPSLPINPGTSHEKH 240
QY 241 EEDGYGFDAIRITAEDESGVMSGHGHNHYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
DB 241 EEDGYGFDAIRITAEDESGVMSGHGHNHYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
QY 301 LSSHEQDYPSPNAKEMKDLKIEKLAGIMQYGVKRESIVVKNKNAIYIPHGDDHHAD 360
DB 301 LSSHEQDYPSPNAKEMKDLKIEKLAGIMQYGVKRESIVVKNKNAIYIPHGDDHHAD 360
QY 361 PIDEHKPVGTGHSNHYELFKPEGVAKKGNKVYTGEEELTNVNLKSTFNQNFLLA 420
DB 361 PIDEHKPVGTGHSNHYELFKPEGVAKKGNKVYTGEEELTNVNLKSTFNQNFLLA 420
QY 421 NGOKRVSFSPPELEKRLGIMLVKLITPDGKYLEKVGKVGEGVGNIANFELDQYLP 480
DB 421 NGOKRVSFSPPELEKRLGIMLVKLITPDGKYLEKVGKVGEGVGNIANFELDQYLP 480
QY 481 GQTFKYTIASKDPEVSYDGTFTVPTSLAYKMASQITFYPFHAGDIYLRVNPQFAPKGT 540
DB 481 GQTFKYTIASKDPEVSYDGTFTVPTSLAYKMASQITFYPFHAGDIYLRVNPQFAPKGT 540
QY 541 DALVRVDEFGHAYLNNYKVGEIKLPIKLNQGTTRAGNKIPVTFMANAYLDNQSY 600
DB 541 DALVRVDEFGHAYLNNYKVGEIKLPIKLNQGTTRAGNKIPVTFMANAYLDNQSY 600
QY 601 IVEVPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSEKVEKELSTGNTSN 660
DB 601 IVEVPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSEKVEKELSTGNTSN 660
QY 661 STLEEVPTVPDQVEKAKFAESYCMKLENVLFNMDGTIELYLPSPGEVVKNNMADTFGEAP 720
DB 661 STLEEVPTVPDQVEKAKFAESYCMKLENVLFNMDGTIELYLPSPGEVVKNNMADTFGEAP 720
QY 721 QGNGENKPSNGKVGSTGTVENQPTENKPADSLPEAPNKPVKPENSTDNGLMPEGNVGS 780
DB 721 QGNGENKPSNGKVGSTGTVENQPTENKPADSLPEAPNKPVKPENSTDNGLMPEGNVGS 780
QY 781 DPMLDLSALEAPAVDPQVEKLEKFTASYGLGLOSVIFNMDGTIELRLPSPGEVVKNNLLIS 840
DB 781 DPMLDLSALEAPAVDPQVEKLEKFTASYGLGLOSVIFNMDGTIELRLPSPGEVVKNNLLIS 840
```

```
DB 781 DPMLDLSALEAPAVDPQVEKLEKFTASYGLGLOSVIFNMDGTIELRLPSPGEVVKNNLLIS 840

RESULT 2
AAU76151
ID AAU76151 standard; Protein: 840 AA.
XX AC AAU76151;
XX DT 08-MAY-2002 (first entry)
XX DE Streptococcus pneumoniae BVH-3 protein version #2.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection.
XX OS Streptococcus pneumoniae.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charla C, Charland N, Martin D, Brodeur B;
XX DR WPI; 2002-122272/16.
XX DR N-PSDB; ABK15105.
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia.
XX PS Example 1; Fig 10; 113pp; English.
XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an
XX CC individual susceptible to these disorders (II) is also useful for
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial
XX CC infection (e.g., caused by Streptococcus pneumoniae, group A
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX CC Staphylococcus aureus) in an individual susceptible to the infection.
XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic
XX CC test for S. pneumoniae infection. (III) is useful for designing DNA
XX CC probes for use in detecting the presence of Streptococcus in a biological
XX CC sample suspected of containing the bacteria. The DNA probes may also be
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for
XX CC diagnosing streptococcal infections. This is the amino acid sequence of
XX CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
XX CC peptides described in the method of the invention.
XX SQ Sequence 840 AA;

Query Match 100.0%; Score 4396; DB 23; Length 840;
Best Local Similarity 100.0%; Pred. No. 1.le-261;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKREGIOAEQIVIKITDQGYV 60
DB 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKREGIOAEQIVIKITDQGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKADIYNEVKGVIIVKDGKYYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKADIYNEVKGVIIVKDGKYYVYLKD 120
```

```
Db 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNEVGKGYIIVKVDGKYVYVLKD 120
Qy 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSOGRTTNDGYVFNPADIIDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSOGRTTNDGYVFNPADIIDTGN 180
Qy 181 YIVPHGHHYHYIPKSDLSASLAAKAHLAKGNMOPQSLSYSTPSPSLPINFGTSHEKH 240
Db 181 YIVPHGHHYHYIPKSDLSASLAAKAHLAKGNMOPQSLSYSTPSPSLPINFGTSHEKH 240
Qy 241 EEDGYGDANRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLAEVKTSHNGLDS 300
Db 241 EEDGYGDANRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLAEVKTSHNGLDS 300
Qy 301 LSSHEQDYPNSAKEMKDLKIEKIAIGIMKQYGVKRESIVVKNKNAIIPHGDDHHAD 360
Db 301 LSSHEQDYPNSAKEMKDLKIEKIAIGIMKQYGVKRESIVVKNKNAIIPHGDDHHAD 360
Qy 361 PIDEHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLA 420
Db 361 PIDEHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLA 420
Qy 421 NGQKRVSFSPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEGVGNIANFELDQYLP 480
Db 421 NGQKRVSFSPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEGVGNIANFELDQYLP 480
Qy 481 GQTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPGKT 540
Db 481 GQTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPGKT 540
Qy 541 DALVRVDFEFGNAYLENNYKVGGEIKLPKPLNOGTRTAGNKTPVTFMANAYLDNOSTY 600
Db 541 DALVRVDFEFGNAYLENNYKVGGEIKLPKPLNOGTRTAGNKTPVTFMANAYLDNOSTY 600
Qy 601 IVEVPILEKENQTDKPSLPOFKRNKAQENSKLDEKVEEPTSEKYEKELSETGNSTSN 660
Db 601 IVEVPILEKENQTDKPSLPOFKRNKAQENSKLDEKVEEPTSEKYEKELSETGNSTSN 660
Qy 661 STLEEVPTVDPQEKVAKFAESYCNKLENVLFNMDGTIELYLPSEVYIKKNMADFTGEAP 720
Db 661 STLEEVPTVDPQEKVAKFAESYCNKLENVLFNMDGTIELYLPSEVYIKKNMADFTGEAP 720
Qy 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGS 780
Db 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGS 780
Qy 781 DPMILDSALEAPADVPQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEVYIKKNLLIS 840
Db 781 DPMILDSALEAPADVPQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEVYIKKNLLIS 840

RESULT 3
AAB12722
ID AAB12722 standard; Protein; 1019 AA.
XX
AC AAB12722;
XX
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO2000039299-A2.
XX
PD 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
```

```
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
XX Claim 18; Fig 20; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae BVH-3M protein antigen.
XX
SQ Sequence 1019 AA:
Query Match 97.3%; Score 4276.5; DB 21; Length 1019;
Best Local Similarity 82.3%; Pred. No. 3.le-254;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;
Qy 1 CAYALNQHSQENKDNRRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIIVIKITDQGVY 60
Db 1 CAYALNQHSQENKDNRRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIIVIKITDQGVY 60
Qy 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNEVGKGYIIVKVDGKYVYVLKD 120
Db 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNEVGKGYIIVKVDGKYVYVLKD 120
Qy 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSOGRTTNDGYVFNPADIIDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSOGRTTNDGYVFNPADIIDTGN 180
Qy 181 YIVPHGHHYHYIPKSDLSASLAAKAHLAKGNMOPQSLSYSTPSPSLPINFGTSHEKH 223
Db 181 YIVPHGHHYHYIPKSDLSASLAAKAHLAKGNMOPQSLSYSTPSPSLPINFGTSHEKH 240
Qy 224 ----- 223
Db 241 PANKSENLOSLLKELYDSPAQRYSSESGLVFDPAKIISRTPNGVAIPHGDHYHFIYSK 300
Qy 224 ----- 223
Db 301 LSALEBKARWVPISGTGSTVSTNAKPNEVWVSSLSNPPSLTTSKELSSASDGYIEN 360
Qy 224 -----TPSPSLPINFGTSHEKHED 243
Db 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQPTLPNNSLATPSLPIPGTSHKHEED 420
Qy 244 GYGFANDRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLAEVKTSHNGLDSLS 303
Db 421 GYGFANDRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLAEVKTSHNGLDSLS 480
Qy 304 HEQDYPNSAKEMKDLKIEKIAIGIMKQYGVKRESIVVKNKNAIIPHGDDHHADPID 363
Db 481 HEQDYPNSAKEMKDLKIEKIAIGIMKQYGVKRESIVVKNKNAIIPHGDDHHADPID 540
Qy 364 EHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLANGQ 423
Db 541 EHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLANGQ 600
Qy 424 KRVSFSPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEGVGNIANFELDQYLPQGT 483
Db 601 KRVSFSPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEGVGNIANFELDQYLPQGT 660
```

```
QY 484 FKYTASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPGKTDAL 543
Db 661 FKYTASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPGKTDAL 720
QY 544 VRVDFEFGHGNAYLENNYKVGKLEIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 603
Db 721 VRVDFEFGHGNAYLENNYKVGKLEIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 780
QY 604 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 663
Db 781 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 840
QY 664 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVIFKKNMADFTGEAPOGN 723
Db 841 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVIFKKNMADFTGEAPOGN 900
QY 724 GENKPSNGKVGSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPM 783
Db 901 GENKPSNGKVGSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPM 960
QY 784 LDSALEEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNL 837
Db 961 LDSALEEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNL 1014

RESULT 4
AAB12751
ID AAB12751 standard; Protein; 1019 AA.
XX AC AAB12751;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN WO2000039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia -
XX PS Disclosure; Fig 11; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the proteins
XX CC antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX CC (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents a
XX CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX SQ Sequence 1019 AA;
```

Query Match 97.3%; Score 4276.5; DB 21; Length 1019;

```
Best Local Similarity 82.3%; Pred. No. 3 le-254;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNHRSQENKDNRRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGVV 60
Db 1 CAYALNHRSQENKDNRRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGVV 60
QY 61 TSHGDHYHYNGVPYDALFESELLMKDPYQLKADIVNEVKGYYLIKVDGKYYVYLD 120
Db 61 TSHGDHYHYNGVPYDALFESELLMKDPYQLKADIVNEVKGYYLIKVDGKYYVYLD 120
QY 121 AAHADNVRTDEINRQKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIETGNA 180
Db 121 AAHADNVRTDEINRQKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIETGNA 180
QY 181 YIYPGHGHYHYIPKSDLSASELAALAAKAGKNNQSQLSYSS----- 223
Db 181 YIYPGHGHYHYIPKSDLSASELAALAAKAGKNNQSQLSYSS----- 223
QY 224 ----- 223
Db 241 PANKSENQSLKELYDPSAQRYSESQGLVDFPAKIISRTPNGVAIPHGDHYHFIPYSK 300
QY 224 ----- 223
Db 301 LSALEEKIARMVPISGTSTVSTNAKNEVYSSLSGSLSSNPSSLTTSKELSSASDGYIFN 360
QY 224 -----TPSPSLPINPGTSHKHEED 243
Db 361 PKDIVETATAYIVRHGDHFHYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHKHEED 420
QY 244 GYGFDANRIIAEDESQFVMSGHGDNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSL 303
Db 421 GYGFDANRIIAEDESQFVMSGHGDNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSL 480
QY 304 HEQDYPNNAKEMKDLKKIEKTAGIMKQYGVKRESIVNVKKNALIIYPHGDHHDADPD 363
Db 481 HEQDYPNNAKEMKDLKKIEKTAGIMKQYGVKRESIVNVKKNALIIYPHGDHHDADPD 540
QY 364 EHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVNLLKNSTFNNQFTLANGQ 423
Db 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVNLLKNSTFNNQFTLANGQ 600
QY 424 KRVSFSPPELEKKGILNMLVKLITPDGKVLKSVKGVGEGVGNIANEELDQYLPQGT 483
Db 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKSVKGVGEGVGNIANEELDQYLPQGT 660
QY 484 FKYTASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPGKTDAL 543
Db 661 FKYTASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPGKTDAL 720
QY 544 VRVDFEFGHGNAYLENNYKVGKLEIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 603
Db 721 VRVDFEFGHGNAYLENNYKVGKLEIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 780
QY 604 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 663
Db 781 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 840
QY 664 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVIFKKNMADFTGEAPOGN 723
Db 841 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVIFKKNMADFTGEAPOGN 900
QY 724 GENKPSNGKVGSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPM 783
Db 901 GENKPSNGKVGSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNGVSDPM 960
QY 784 LDSALEEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNL 837
Db 961 LDSALEEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNL 1014

RESULT 5
```


AAU84021

ID AAU84021 standard; Peptide; 1019 AA.

XX AAU84021;

XX 08-MAY-2002 (first entry)

DE Truncated variant of *S. pneumoniae* BVH-3, BVH-3M.XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.

XX Streptococcus pneumoniae.

OS Synthetic.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Example 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine of comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for *S. pneumoniae* infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating *S. pneumoniae* nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX Sequence 1019 AA;

Query Match 97.3%; Score 4276.5; DB 23; Length 1019;

Best Local Similarity 82.3%; Pred. No. 3.le-254;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQRSEKDNRRSYVDGSSQSKSENTPQVSKQEGIAEQIVIKITDGGYV 60

DB 1 CAYALNQRSEKDNRRSYVDGSSQSKSENTPQVSKQEGIAEQIVIKITDGGYV 60

QY 61 TSHGDHYHYNGKPYDALFSEELMKDPNVLKADIVNEVGKGYIIVKDGKYYVVLKD 120

DB 61 TSHGDHYHYNGKPYDALFSEELMKDPNVLKADIVNEVGKGYIIVKDGKYYVVLKD 120

QY 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRVTTNDGYVFNPAIDTIGNA 180
DB 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRVTTNDGYVFNPAIDTIGNA 180
QY 181 YIVPHGGHYHYIPKSDLSASELAATAKAHLAGKNMQPSQLSYSS- 223
DB 181 YIVPHGGHYHYIPKSDLSASELAATAKAHLAGKNMQPSQLSYSS- 240
QY 224 - 223
DB 241 PANKSENLOSLLKELYDSPAQRYSESGLVFPDPAKIIISRTPNGVAIPHGDHYHFPYK 300
QY 224 - 223
DB 301 LSALEEKIARMVPISGTGVSTVSTNAKNEVWSSGLSSNPPSLTTSKELSSASDGYIEN 360
QY 224 - 223
DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 244 GYGFDANRIIAEDESQFVMSHGDHNYHFFKKDLTEQIKAAQKHLEVKTSNGLDLSLS 303
DB 421 GYGFDANRIIAEDESQFVMSHGDHNYHFFKKDLTEQIKAAQKHLEVKTSNGLDLSLS 480
QY 304 HEQDYPNNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIYPHGDHHAADPID 363
DB 481 HEQDYPNNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIYPHGDHHAADPID 540
QY 364 EHKPVGIGHSHSNYELFKPEGVAKKEGKNVYTGEBLTNVNLLKNSTNNQNTLANQ 423
DB 541 EHKPVGIGHSHSNYELFKPEGVAKKEGKNVYTGEBLTNVNLLKNSTNNQNTLANQ 600
QY 424 KRVSFSPFPELEKKLGINMLVKLITPDGKVLKVSQKFGEGVGNIANFELDQPYLPQGT 483
DB 601 KRVSFSPFPELEKKLGINMLVKLITPDGKVLKVSQKFGEGVGNIANFELDQPYLPQGT 660
QY 484 FKVTIASKQDPEVSYDGTFTVPTSLAYKASQTIFFYFFHAGDTYLRVNPQFAVPKGTDAL 543
DB 661 FKVTIASKQDPEVSYDGTFTVPTSLAYKASQTIFFYFFHAGDTYLRVNPQFAVPKGTDAL 720
QY 544 VRVDFEFGHNAVLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTETMANAYLDNQSYIVE 603
DB 721 VRVDFEFGHNAVLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTETMANAYLDNQSYIVE 780
QY 604 VPILENQTDKPSIILPQKRKAQNSKLDKVEEPKTSKVEKEKLSGTGNSNSTL 663
DB 781 VPILENQTDKPSIILPQKRKAQNSKLDKVEEPKTSKVEKEKLSGTGNSNSTL 840
QY 664 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKNNMADFTGEAPOGN 723
DB 841 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKNNMADFTGEAPOGN 900
QY 724 GENKPSNGKVSTGVENOPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNVGSDBM 783
DB 901 GENKPSNGKVSTGVENOPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNVGSDBM 960
QY 784 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPDSGEVKKNL 837
DB 961 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPDSGEVKKNL 1014

RESULT 6

AAB12715

ID AAB12715 standard; Protein; 1039 AA.

XX AAB12715;

AC AAB12715;

DT 21-NOV-2000 (first entry)

XX

Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.

DE

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

N-PSDB; AAA65730.

Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteraemia and/or pneumonia

Claim 18; Fig 2; 106pp; English.

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3 protein antigen.

Sequence 1039 AA;

Query Match 97.3%; Score 4276.5; DB 21; Length 1039;

Best Local Similarity 82.3%; Pred. No. 3.1e-254; Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITDQGIY 60
21 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITDQGIY 80
61 TSHGDHYHYNGKVPYDALFSEELMKPNYOLKADIVNEVKGGYIIKVDCKYVYLKD 120
81 TSHGDHYHYNGKVPYDALFSEELMKPNYOLKADIVNEVKGGYIIKVDCKYVYLKD 140
121 AAHADNVRTKDEINRQKQEHVKNKNSNVAARSQGRYTTNDGYVFPADIIEDTGN 180
141 AAHADNVRTKDEINRQKQEHVKNKNSNVAARSQGRYTTNDGYVFPADIIEDTGN 200
181 YIVPHGGHYHYTPKSDLSASELAHAKAHLAGKNMOPSOISYSS----- 223
201 YIVPHGGHYHYTPKSDLSASELAHAKAHLAGKNMOPSOISYSSASDNNQTSVAKGSTSK 260
224 ----- 223
261 PANKSENLOSLLKELYDPSAQYSESGLVDPKAKIISRTPNGVAIPHGDHYHPIYSK 320
224 ----- 223
321 LSALAEKIAARMVPISGTGVSTNAKPNNEWSSGLSSNPSLTSKELSSASDGIYN 380
224 -----TPSPSLPINPGTSHKHEED 243
381 PKDIVETATATVIRGDFHVIYIPKSNQIQPTLNNSLATPSPSLPINPGTSHKHEED 440
244 GYGFDAIRIIAEDSEGFVMSHGDHNNHYFFKDLTEEQKAAQKHLSEEVKTSNGLDLS 303
441 GYGFDAIRIIAEDSEGFVMSHGDHNNHYFFKDLTEEQKAAQKHLSEEVKTSNGLDLS 500
304 HEQDYPSNAKEMKOLDKTEEKIAGTMQYGVKRESIVVKNKNAIYPHGDHHDADPID 363

Db 501 HEQDYPSNAKEMKOLDKTEEKIAGTMQYGVKRESIVVKNKNAIYPHGDHHDADPID 560
QY 364 EHKPVGIGHSNYSYELFKPEEGVAKKGGKVVYTGELTNVNLKKNSTNNQNFLLANQ 423
Db 561 EHKPVGIGHSNYSYELFKPEEGVAKKGGKVVYTGELTNVNLKKNSTNNQNFLLANQ 620
QY 424 KRVSFPPPELEKKGINMLVLIITPDGKVLKVSQKVFEGVGNIANFELDQVLPQGT 483
Db 621 KRVSFPPPELEKKGINMLVLIITPDGKVLKVSQKVFEGVGNIANFELDQVLPQGT 680
QY 484 FKTIASKOYPEVSYDGTFTVPTSLAYKASQTIFFPFHAGDTYLRVNPQFAVPKGTDL 543
Db 681 FKTIASKOYPEVSYDGTFTVPTSLAYKASQTIFFPFHAGDTYLRVNPQFAVPKGTDL 740
QY 544 VRFDEFHGNAYLNNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
Db 741 VRFDEFHGNAYLNNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 800
QY 604 VPILKENQTDKPSILPQFKRNKAQENSKLDERVEPKTSEKVEKEKLSNLSN 563
Db 801 VPILKENQTDKPSILPQFKRNKAQENSKLDERVEPKTSEKVEKEKLSNLSN 860
QY 564 EEPVTVDPQVEKVAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 723
Db 861 EEPVTVDPQVEKVAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 920
QY 724 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPM 783
Db 921 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSDPM 980
QY 784 LDSALEAPADVPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSEVTKKNL 837
Db 981 LDSALEAPADVPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSEVTKKNL 1034

RESULT 7

AAU75932
ID AAU75932 standard; Protein; 1039 AA.

XX AAU75932;

DT 08-MAY-2002 (first entry)

XX Streptococcus pneumoniae BVH-3 protein version #1.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection.

XX Streptococcus pneumoniae.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/15.

XX N-PSDB; ABK15101.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Example 1; Fig 6; 113pp; English.

XX

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This is the amino acid sequence of
CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
CC peptides described in the method of the invention.
xx
SQ Sequence 1039 AA;

Query Match 97.3%; Score 4276.5; DB 23; Length 1039;
Best Local Similarity 82.3%; Pred. No. 3.1e-254;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQRHSEKNNRVSVYDGSQSSOKSENTPDOVSQKEGIAQIIVIKITDQGYV 60
DB 21 CAYALNQRHSEKNNRVSVYDGSQSSOKSENTPDOVSQKEGIAQIIVIKITDQGYV 80
QY 61 TSHGDPHYHYNGKVPYDALFSEELMKDPNQLKDADIVNEVKGGYIIVKDGKYYVVLKD 120
DB 81 TSHGDPHYHYNGKVPYDALFSEELMKDPNQLKDADIVNEVKGGYIIVKDGKYYVVLKD 140
QY 121 RAHADNVRTKEINRQKQEHVKNKNSVAVARQSQRYTNDGYVFNPAIDIEDTGN 180
DB 141 RAHADNVRTKEINRQKQEHVKNKNSVAVARQSQRYTNDGYVFNPAIDIEDTGN 200
QY 181 YIVPHGGHYHYIPKSDLSASLAAKAHLAKGNMQPSQSYSS----- 223
DB 201 YIVPHGGHYHYIPKSDLSASLAAKAHLAKGNMQPSQSYSSSTASDNNQTSVAKGSTSK 260
QY 224 ----- 223
DB 261 PANKSENLOLLKELYDSPAQRVSESDGLVDPAPAKIISRTPNGVAPLPHGDHYHFIYSK 320
QY 224 ----- 223
DB 321 LSALPEKIARWPIISGTGSTVSTNAKNPNEVSVSLGSLSSNPSSLTTSKELSSASDGVIFN 380
QY 224 -----TPSPSLPINPCTSHKHEED 243
DB 381 PKDIVVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLPINPCTSHKHEED 440
QY 244 GYGFDANRIIAEDSGFVMSGHDHNYEFKKDLTEEQKAAQKHLEEVKTSNHLGDSLS 303
DB 441 GYGFDANRIIAEDSGFVMSGHDHNYEFKKDLTEEQKAAQKHLEEVKTSNHLGDSLS 500
QY 304 HQDYPGNAKEMKDLKDKIEEKIAGIMKQYGVKRESIVWNEKNAIYPHGDHHAAPID 363
DB 501 HQDYPGNAKEMKDLKDKIEEKIAGIMKQYGVKRESIVWNEKNAIYPHGDHHAAPID 560
QY 364 EHKPVGIGHSHSNVLELFPPEGVAKKNGKYYTGEELTNVNLKNSFTNQNFTLANGQ 423
DB 561 EHKPVGIGHSHSNVLELFPPEGVAKKNGKYYTGEELTNVNLKNSFTNQNFTLANGQ 620
QY 424 KRVSFPFPELEKILGINMLVKLITPDGKVLKYSKGVFGVGNIANFELDQPYLPQGT 483
DB 621 KRVSFPFPELEKILGINMLVKLITPDGKVLKYSKGVFGVGNIANFELDQPYLPQGT 680
QY 484 FKYTITASKDPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 543

DB 681 FKYTITASKDPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 740
QY 544 VRVDEEHGDAYLENNYKVGIEKLPIEKLKNOGTRTAGNKIPVTFEMANAYLDNQSYIVE 603
DB 741 VRVDEEHGDAYLENNYKVGIEKLPIEKLKNOGTRTAGNKIPVTFEMANAYLDNQSYIVE 800
QY 504 VPILEKENQTDKPSILPQFRKRAQENSKLDEKVEEPKTSKEVKEKELSETGNSNSTL 663
DB 801 VPILEKENQTDKPSILPQFRKRAQENSKLDEKVEEPKTSKEVKEKELSETGNSNSTL 860
QY 664 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGVIRKKNMADFTGEAPQGN 723
DB 861 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGVIRKKNMADFTGEAPQGN 920
QY 724 GENKPSNGKVGSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLMPEGNVGSDDPM 783
DB 921 GENKPSNGKVGSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLMPEGNVGSDDPM 980
QY 784 LDSALEAPAVDPVOEKLKFTASYGLGLDSVIFNMDGTIELRPLPSGEVIRKKNL 837
DB 981 LDPALEAPAVDPVOEKLKFTASYGLGLDSVIFNMDGTIELRPLPSGEVIRKKNL 1034

RESULT 8
AAB12750
ID AAB12750 standard; Protein; 1019 AA.
XX
AC AAB12750;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN W0200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
PS Disclosure; Fig 11; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
SQ Sequence 1019 AA;

Query Match 97.2%; Score 4274.5; DB 21; Length 1019;
Best Local Similarity 82.3%; Pred. No. 4.1e-254;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

```
QY 1 CAYALNHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGV 60
DB 1 CAYALNHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGGYIIKVDGKYVYVKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGGYIIKVDGKYVYVKD 120
QY 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSOGRYTTNDGYVFNPAIDTGTNA 180
DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSOGRYTTNDGYVFNPAIDTGTNA 180
QY 181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSS 223
DB 181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSS 223
QY 241 PANKSENLOSLLKELYDPSAQRYSESGLVDPAPAKIISRTPNGVAIPHGDHYHFIYSK 300
DB 241 PANKSENLOSLLKELYDPSAQRYSESGLVDPAPAKIISRTPNGVAIPHGDHYHFIYSK 300
QY 301 LSALEEKIARMVPISGTGVSTNAKPNEVSSLGSLSSNPSSLTTSKELSSASDGYIFN 360
DB 301 LSALEEKIARMVPISGTGVSTNAKPNEVSSLGSLSSNPSSLTTSKELSSASDGYIFN 360
QY 224 -----TPSPSLPINPGTSHKHEED 243
DB 224 -----TPSPSLPINPGTSHKHEED 243
QY 361 PKDIVEETATAVIRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
DB 361 PKDIVEETATAVIRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
QY 244 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 303
DB 244 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 303
QY 421 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 480
DB 421 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 480
QY 304 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 363
DB 304 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 363
QY 481 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 540
DB 481 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 540
QY 364 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 423
DB 364 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 423
QY 541 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 600
DB 541 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 600
QY 424 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
DB 424 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
QY 601 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 660
DB 601 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 660
QY 484 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 543
DB 484 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 543
QY 661 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 720
DB 661 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 720
QY 544 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
DB 544 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
QY 721 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
DB 721 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
QY 604 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 663
DB 604 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 663
QY 781 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 840
DB 781 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 840
QY 664 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 723
DB 664 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 723
QY 841 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 900
DB 841 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 900
QY 724 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 783
DB 724 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 783
QY 901 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 960
DB 901 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 960
QY 784 LSALEEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 837
DB 784 LSALEEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 837
QY 961 LDPALAEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 1014
DB 961 LDPALAEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 1014
RESULT 9
AAB12748
ID AAB12748 standard; Protein: 1019 AA.
```

XX AAB12748:
XX AC
XX 21-NOV-2000 (first entry)
XX DE
XX Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
XX DE
XX Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal.
XX OS
XX Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia --
XX PS Disclosure; Fig 11; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein
XX CC antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX CC (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents a
XX CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX SQ Sequence 1019 AA:
Query Match 97.1%; Score 4268.5; DB 21; Length 1019;
Best Local Similarity 82.2%; Pred. No. 9.5e-254;
Matches 834; Conservative 0; Mismatches 3; Indels 177; Gaps 1:
QY 1 CAYALNHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGV 60
DB 1 CAYALNHRQENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQAEQIVIKITDQGV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGGYIIKVDGKYVYVKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGGYIIKVDGKYVYVKD 120
QY 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSOGRYTTNDGYVFNPAIDTGTNA 180
DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSOGRYTTNDGYVFNPAIDTGTNA 180
QY 181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSS 223
DB 181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMOPSOLESYSS 223
QY 224 -----TPSPSLPINPGTSHKHEED 243
DB 224 -----TPSPSLPINPGTSHKHEED 243
QY 241 PANKSENLOSLLKELYDPSAQRYSESGLVDPAPAKIISRTPNGVAIPHGDHYHFIYSK 300
DB 241 PANKSENLOSLLKELYDPSAQRYSESGLVDPAPAKIISRTPNGVAIPHGDHYHFIYSK 300
QY 301 LSALEEKIARMVPISGTGVSTNAKPNEVSSLGSLSSNPSSLTTSKELSSASDGYIFN 360
DB 301 LSALEEKIARMVPISGTGVSTNAKPNEVSSLGSLSSNPSSLTTSKELSSASDGYIFN 360
QY 224 -----TPSPSLPINPGTSHKHEED 243
DB 224 -----TPSPSLPINPGTSHKHEED 243
QY 361 PKDIVEETATAVIRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
DB 361 PKDIVEETATAVIRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
QY 244 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 303
DB 244 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 303
QY 421 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 480
DB 421 GYCFDANRIIADESGFVMSHGDHNYFFKDLTEEQIAKQKHLEEVKTSNGLDLS 480
QY 304 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 363
DB 304 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 363
QY 481 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 540
DB 481 HQDTPSNKAKMDLKKTEEKIAGIKQYGVKRESIVVNEKNAIIPHGDDHHDADID 540
QY 364 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 423
DB 364 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 423
QY 541 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 600
DB 541 EKPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNVNVLLKNSFNQNF LANGQ 600
QY 424 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
DB 424 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
QY 601 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 660
DB 601 KRVSFSPPELEKKLGINMLVLIPTDQKVKLEKSGKVGEGVGNIANFELDQYLPQGT 660
QY 484 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 543
DB 484 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 543
QY 661 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 720
DB 661 FKYTIAKQDYPEVSYDGTFTVPTSLAYKASOTIFYPHAGDTYLRVNPQFAVPKGTAL 720
QY 544 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
DB 544 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
QY 721 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
DB 721 VRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
QY 604 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 663
DB 604 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 663
QY 781 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 840
DB 781 VPILEKENTDPSILPQFKRKAQENKLDKVEEPTSEKVEKEKLESETGNSNSTL 840
QY 664 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 723
DB 664 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 723
QY 841 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 900
DB 841 EEPVTVDPQVQKFAEYSYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPOGN 900
QY 724 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 783
DB 724 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 783
QY 901 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 960
DB 901 GENKPSNGKSVGTGVNPTENKPADSLPEAPNEKPKVPENSTONGMLNPGVNGSDPM 960
QY 784 LSALEEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 837
DB 784 LSALEEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 837
QY 961 LDPALAEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 1014
DB 961 LDPALAEAPAVDPQVQKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNL 1014
RESULT 9
AAB12748
ID AAB12748 standard; Protein: 1019 AA.

Db 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQGPTLPNNSLATPSPLPINPGTSHKHEED 420
 QY 244 GYGFDANRIIAEDESFGVMSHGDHNNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 303
 Db 421 GYGFDANRIIAEDESFGVMSHGDHNNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 480
 QY 304 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 363
 Db 481 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 540
 QY 364 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 423
 Db 541 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 600
 QY 424 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 483
 Db 601 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 660
 QY 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 543
 Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 720
 QY 544 VRVDFEFGHGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603
 Db 721 VRVDFEFGHGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
 QY 604 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEEPKTSKVEKEKLSSETGNSTNSL 663
 Db 781 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEEPKTSKVEKEKLSSETGNSTNSL 840
 QY 664 EYVPTVDPVQEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVIKKNMADFTGEAPQGN 723
 Db 841 EYVPTVDPVQEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVIKKNMADFTGEAPQGN 900
 QY 724 GENKPSGKNGSVGTVENOPTENKPADSLPEAPNEXPKPENSTDNGLNPNPENGVSDEPM 783
 Db 901 GENKPSGKNGSVGTVENOPTENKPADSLPEAPNEXPKPENSTDNGLNPNPENGVSDEPM 960
 QY 784 LDSALEEAPADVPQVKLEKFTASYGLGLDSVIFNMMDGTIELRPSGEVKKNL 837
 Db 961 LDPALAEAPADVPQVKLEKFTASYGLGLDSVIFNMMDGTIELRPSGEVKKNL 1014

RESULT 10

AAB12752
 ID AAB12752 standard; Protein; 1019 AA.

AC AAB12752;

XX 21-NOV-2000 (first entry)

DT Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.

DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine.
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

OS WO200039299-A2.

PN 06-JUL-2000.

PD 20-DEC-1999; 99WO-CA01218.

PF 23-DEC-1998; 98US-0113800.

PR (BIOC-) BIOCHEM PHARMA INC.

PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

PI WPI; 2000-452397/39.

XX

XX Streptococcal proteins useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia
 XX Disclosure; Fig 11; 106pp; English.
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX Sequence 1019 AA;

Query Match 97.1%; Score 4268.5; DB 21; Length 1019;
 Best Local Similarity 82.2%; Pred. No. 9.5e-254;

Matches 834; Conservative 0; Mismatches 3; Indels 177; Gaps 1;

QY 1 CAYALNQRHQENKNNRVSYVDGSSQSKSENLTDPDVSQKEGIAEQIVIKITDQYV 60

Db 1 CAYALNQRHQENKNNRVSYVDGSSQSKSENLTDPDVSQKEGIAEQIVIKITDQYV 60

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLDADIVNEVKGYIITKVDGKYVYVUKD 120

Db 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLDADIVNEVKGYIITKVDGKYVYVUKD 120

QY 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQRYTTNDGYVFNPAIIEEDTGN 180

Db 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQRYTTNDGYVFNPAIIEEDTGN 180

QY 181 YIVPHGGHYHYTPKSDLSASELAHAKHAGKNMPSLSYSS 223

Db 181 YIVPHGGHYHYTPKSDLSASELAHAKHAGKNMPSLSYSS 223

QY 241 PANKSENQLSKLLELYDSPAQRYSQSDGLVDPAPAKIISRTPNGVAIPHGDIYHFIPYSK 300

Db 241 PANKSENQLSKLLELYDSPAQRYSQSDGLVDPAPAKIISRTPNGVAIPHGDIYHFIPYSK 300

QY 224 223

Db 224 223

QY 301 LSALKEEIKARMVPISGTGVSTNAKPREVSSLSGLSSNPSLSLTTSKELSSASDGYIFN 360

Db 301 LSALKEEIKARMVPISGTGVSTNAKPREVSSLSGLSSNPSLSLTTSKELSSASDGYIFN 360

QY 224 243

Db 224 243

QY 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQGPTLPNNSLATPSPLPINPGTSHKHEED 420

Db 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQGPTLPNNSLATPSPLPINPGTSHKHEED 420

QY 244 GYGFDANRIIAEDESFGVMSHGDHNNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 303

Db 244 GYGFDANRIIAEDESFGVMSHGDHNNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 303

QY 304 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 363

Db 304 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 363

QY 481 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 540

Db 481 HEODYPSNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVVNKEKNALIIYPHGDHHAADPID 540

QY 364 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 423

Db 364 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 423

QY 541 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 600

Db 541 EHPVGIGSHSHSYELFKPEEGVAKKGEKNKVTGEBELTNVNLKNSTFNNQNTFLANGQ 600

QY 424 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 483

Db 424 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 483

QY 601 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 660

Db 601 KRVSFPPEPELEKLGINMLVLIITPDGKVLKESKVKFGEGVGNIANFELDQPYLPQGT 660

QY 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 543

Db 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 543

QY 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 720

Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPKGTDAL 720

QY 544 VRVDFEFGHGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603

Db 544 VRVDFEFGHGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603

Db 721 VRVDFEFGHGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780

```
QY 604 VPILKENTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSSTL 563
DB 781 VPILKENTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSSTL 840
QY 664 EEPVTPDVPQEKVAFESYGMKLENVLFNMDGTIELYLPSEGEVKKKNAADFTGEAPOGN 723
DB 841 EEPVTPDVPQEKVAFESYGMKLENVLFNMDGTIELYLPSEGEVKKKNAADFTGEAPOGN 900
QY 724 GENKPSENGKSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNVGSDFM 783
DB 901 GENKPSENGKSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNVGSDFM 960
QY 784 LSALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 837
DB 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 1014

RESULT 11
AAB12753
ID AAB12753 standard; Protein; 1019 AA.
XX AC AAB12753;
XX AC AAB12753;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae strain A66 BVH-3 protein antigen.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX ST Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia.
XX PS Disclosure; Fig 11; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein
XX CC antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX CC (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents a
XX CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX SQ Sequence 1019 AA;
Query Match 97.1%; Score 4268.5; DB 21; Length 1019;
Best Local Similarity 82.2%; Pred. No. 9.5e-254;
Matches 834; Conservative 0; Mismatches 3; Indels 177; Gaps 1;
QY 1 CAYALNQHRSQENKDNRRVSYVDGSSQSKSENLTPOQVSKQEGIAEQIVIKITDQGYV 60
DB 1 CAYALNQHRSQENKDNRRVSYVDGSSQSKSENLTPOQVSKQEGIAEQIVIKITDQGYV 60
```

```
QY 61 TSHGDHYHYNGKVPYDALPSEELLKMKDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALPSEELLKMKDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLKD 120
QY 121 AAHADNVRTKDEINRQKQEHVKONEKNSVNAVARSOGRTYTTNDGIVFNPADIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKQEHVKONEKNSVNAVARSOGRTYTTNDGIVFNPADIIEDTGN 180
QY 181 YIVPHGCHYHYIPKSDLSASELAAKHAHLAGKNQPSOLSYS----- 223
DB 181 YIVPHRCHYHYIPKSDLSASELAAKHAHLAGKNQPSOLSYSSTASDNNTQSVAKGSTK 240
QY 224 ----- 223
DB 241 PANKSENQLSKELYDPSAQRYSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
QY 224 ----- 223
DB 301 LSALAEKIARWPIISGTGSTVSTNAKNPNEVVSSLSGSLSSNPSSLTTSKELSSASDGYIFN 360
QY 224 -----TPSPSLPINPGETSHEKHEED 243
DB 361 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPINPGETSHEKHEED 420
QY 244 GYGFDA NR IIAEDSGFVMSHGDNHNYFFKKDLTEBOIKAAQKHLEEVKTSNGLDSLSS 303
DB 421 GYGFDA NR IIAEDSGFVMSHGDNHNYFFKKDLTEBOIKAAQKHLEEVKTSNGLDSLSS 480
QY 304 HEQYPSNAKEMKOLDKKIEBKIAQKQYGVKRESIVNKEKNALIIYPHGDHHDADPID 363
DB 481 HEQYPSNAKEMKOLDKKIEBKIAQKQYGVKRESIVNKEKNALIIYPHGDHHDADPID 540
QY 364 EHKPVGICHSHSNYELFKPPEGVAKKEGNKVYTGEELTNVNNLLKNSTFNQNFILANGQ 423
DB 541 EHKPVGICHSHSNYELFKPPEGVAKKEGNKVYTGEELTNVNNLLKNSTFNQNFILANGQ 600
QY 424 KRVSFPPELEKLGINLVKLTTPDGKVLKESGVGEGVGNIANFELDQYLPQGQT 483
DB 601 KRVSFPPELEKLGINLVKLTTPDGKVLKESGVGEGVGNIANFELDQYLPQGQT 660
QY 484 FKYTIASKDYPEVSYDGTFTVPTSLAYKNASQTIYFPHAGDTYLRVNPQFAVPGKTDAL 543
DB 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKNASQTIYFPHAGDTYLRVNPQFAVPGKTDAL 720
QY 544 VEVDFEFHGNAYLENNYKVGTEIKLPIPKLNOGTTTACNKIPVTFMANAYLDNOSTYIVE 603
DB 721 VRVDFEFHGNAYLENNYKVGTEIKLPIPKLNOGTTTACNKIPVTFMANAYLDNOSTYIVE 780
QY 604 VPILKENTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSSTL 663
DB 781 VPILKENTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSSTL 840
QY 664 EEPVTPDVPQEKVAFESYGMKLENVLFNMDGTIELYLPSEGEVKKKNAADFTGEAPOGN 723
DB 841 EEPVTPDVPQEKVAFESYGMKLENVLFNMDGTIELYLPSEGEVKKKNAADFTGEAPOGN 900
QY 724 GENKPSENGKSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNVGSDFM 783
DB 901 GENKPSENGKSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNVGSDFM 960
QY 784 LSALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 837
DB 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 1014

RESULT 12
AAB12749
ID AAB12749 standard; Protein; 1019 AA.
XX AC AAB12749;
XX AC AAB12749;
XX DT 21-NOV-2000 (first entry)
XX DT 21-NOV-2000 (first entry)
XX DT 21-NOV-2000 (first entry)
```

DE Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia
 XX
 PS Disclosure; Fig 11; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SQ Sequence 1019 AA;

Query Match 97.08; Score 4264.5; DB 21; Length 1019;
 Best Local Similarity 82.18; Pred. No. 1.7e-253;
 Matches 832; Conservative 2; Mismatches 3; Indels 177; Gaps 1;
 QY 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTPOVSKQEGTQAEQIVIKITDQGYV 60
 DB 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTPOVSKQEGTQAEQIVIKITDQGYV 60
 QY 61 TSHGDHYHYNGKPYDPLFSEELLMDPNYQLKDADIVNEVKGGYTIKVDGKYVYVLDK 120
 DB 61 TSHGDHYHYNGKPYDPLFSEELLMDPNYQLKDADIVNEVKGGYTIKVDGKYVYVLDK 120
 QY 121 AAHADNVRTKDEINRQKQEHVKDKENKNSVAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
 DB 121 AAHADNVRTKDEINRQKQEHVKDKENKNSVAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
 QY 181 YIVPHGGHYHYIPKSDLSASLAAKAHLAGKNMOPQSOLSYSS----- 223
 DB 181 YIVPHGGHYHYIPKSDLSASLAAKAHLAGKNMOPQSOLSYSSSTASDNTOSVAKGSTK 240
 QY 224 ----- 223
 DB 241 PANKSENLOSLLKELYDSPSAQRYSSEGLVDPDAKIISRTPNGVAIPHGDHYHFIPYSK 300
 QY 224 ----- 223
 DB 301 LSALBEKIARRVPISTGTSTVNAKPNEVVSSLSLSSNPSSLTTSKELSSASDGYIFN 360
 QY 224 -----TPSPSLINPOTSHEKHEED 243
 DB 361 PKDIVEETATYIVRHGDHYHYIPKSNQIGOPTLPNNSLATPSPSLPINPOISHEKHEED 420
 QY 244 GYGF DANRIIAEDSGFIMSHGNHNYFFKKDLTTEQIKAAQKHLEEVKTSNGLDLSLS 303
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia

DB 421 GYGF DANRIIAEDSGFIMSHGNHNYFFKKDLTTEQIKAAQKHLEEVKTSNGLDLSLS 480
 QY 304 HEQDYPNANEMKMDLKKIEEKTAGINKQYGVKRESIVVKNKNALIIYPGDDHHHADPID 363
 DB 481 HEQDYPNANEMKMDLKKIEEKTAGINKQYGVKRESIVVKNKNALIIYPGDDHHHADPID 540
 QY 364 EHKPVGIGCHSHSNYELFKPEEGYAKKKNKVVYTGEEELTNVNNLKNSTENNNQFTLANGQ 423
 DB 541 EHKPVGIGCHSHSNYELFKPEEGYAKKKNKVVYTGEEELTNVNNLKNSTENNNQFTLANGQ 600
 QY 424 KRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSCKVFGEGVGNANTANFELDQYLPQGT 483
 DB 601 KRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSCKVFGEGVGNANTANFELDQYLPQGT 660
 QY 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAYPKGTDAL 543
 DB 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAYPKGTDAL 720
 QY 544 VRFDFPHGNAYLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 603
 DB 721 VRFDFPHGNAYLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780
 QY 604 VPILKEKNQDKPSILPOFKRNKAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSSTL 663
 DB 781 VPILKEKNQDKPSILPOFKRNKAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSSTL 840
 QY 664 EEVPTVPDQVEKVAFAESYGMKLENVLFNMDDGTIELYLPSEGVIRKKNMADFTGEAPQGN 723
 DB 841 EEVPTVPDQVEKVAFAESYGMKLENVLFNMDDGTIELYLPSEGVIRKKNMADFTGEAPQGN 900
 QY 724 GENKPSNGKSVGTGTVENQGTENKPADSLPEAPNEKPVKPESTDNQMLNPEGNVGSDDPM 783
 DB 901 GENKPSNGKSVGTGTVENQGTENKPADSLPEAPNEKPVKPESTDNQMLNPEGNVGSDDPM 960
 QY 784 LDSALEAPAVDPVQEKLEFTASYGLGDSVIFNMDGTIELRLPSEGVIRKKNL 837
 DB 961 LDSALEAPAVDPVQEKLEFTASYGLGDSVIFNMDGTIELRLPSEGVIRKKNL 1014
 RESULT 13
 AAU83997
 ID AAU83997 standard; Peptide; 807 AA.
 XX
 AC AAU83997;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW25.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia

XX PS Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX CC

XX Sequence 807 AA;

Query Match 77.4%; Score 3402.5; DB 23; Length 807;
 Best Local Similarity 79.6%; Pred. No. 1.1e-200;
 Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11:

QY 10 SOENKDNRRVYVYDGSQS--SKSENLL-----TPDQVSKQEGTQAEQI-VIKIT 55
 DB 10 SSTASDNTQSVAKGSTSKPANKSENLLKELYDPSAQRYSESGLVDFDPAKIISRT 69
 QY 56 DOGYVTSHGHTHYNGKVPYDALPS-ELLMKDPNYQLKDADIVNEVKGVIIVKVDGY 114
 DB 70 PNGVAIPGHGDHYF-----IPYKLSALEEKIAR-----MVPISGT- 105
 QY 115 YVYLKDAHADNRVTKDEINRQKHVKDNKVNMSVAVARSQGRYTT-----NDGYV 167
 DB 106 -----GSTVST-----NAKNEVVSILGSLSSNPSSILTSKELSSASDGYI 146
 QY 168 FNPADIIETGNAYIVPHGCHVHYTPKSDLSASELAAKAHLAGNMOPSQLSYS-STPS 226
 DB 147 FNPKDIVEETATAYIVRHGDHFHYTPKSNQIG-----OPTLFNNSLATPS 191
 QY 227 PSLPIMPNTSHEKHEDGFGFDANRTIAEDSGFVMSHGDHNYFFKDLTEEQIKAQK 286
 DB 192 PSLPTIPNGTSHEKHEDGFGFDANRTIAEDSGFVMSHGDHNYFFKDLTEEQIKAQK 251
 QY 287 HLEVKTSHNGLDLSSSHEDQYPSNAKEMKDLKKEEKIAGIMQYGVKRESIVVYVNEK 346
 DB 252 HLEVKTSHNGLDLSSSHEDQYPSNAKEMKDLKKEEKIAGIMQYGVKRESIVVYVNEK 311
 QY 347 NAIYPHGDHHDADPDEKHPVIGHSHSNYELFPPEGVAKKEGKNVYTGELTNVNL 406
 DB 312 NAIYPHGDHHDADPDEKHPVIGHSHSNYELFPPEGVAKKEGKNVYTGELTNVNL 371
 QY 407 LKNSTNNONFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKVLKESGKVFGEV 466
 DB 372 LKNSTNNONFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKVLKESGKVFGEV 431
 QY 467 GNTANPELOQPLPGGTFFYTTASKDYPVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 526
 DB 432 GNTANPELOQPLPGGTFFYTTASKDYPVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 491
 QY 527 YLRVNPQFAVPKGTDALVRFDFEFGNAYLNNYKVGELKIPKLNOCGTTTAGNKIPV 586
 DB 492 YLRVNPQFAVPKGTDALVRFDFEFGNAYLNNYKVGELKIPKLNOCGTTTAGNKIPV 551
 QY 587 TFMANAYLNQSTYIVVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 646

DB 552 TFMANAYLNQSTYIVVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 611
 QY 647 EKEKLSGTGNTSNTLEEVPTVDPVQEKVAKFAESVGMKLENVLFNMDGTIELYLPSPGE 706
 DB 612 EKEKLSGTGNTSNTLEEVPTVDPVQEKVAKFAESVGMKLENVLFNMDGTIELYLPSPGE 671
 QY 707 VIKKNMADFTGEAPQNGENKPSFENGKSVGTVENOPTENKPADSLPEAPNEKPVKAPENS 766
 DB 672 VIKKNMADFTGEAPQNGENKPSFENGKSVGTVENOPTENKPADSLPEAPNEKPVKAPENS 731
 QY 767 TDGMLNPEGVSDPMDLSALEEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELR 826
 DB 732 TDGMLNPEGVSDPMDLPALDEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELR 791
 QY 827 LPSGEVIRKKNL 837
 DB 792 LPSGEVIRKKNL 802

RESULT 14
 AAU84054
 ID AU84054 standard; Peptide: 1152 AA.
 XX AC AU84054;
 XX DT 08-MAY-2002 (first entry)
 XX DE S. pneumoniae derived chimeric peptide, NEW27.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; muten;
 KW BVH-11-2.
 XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX PN WO200198334-A2.
 XX PD 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 XX 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 WPI; 2002-122272/16.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 epitope-bearing polypeptides, useful as vaccine components for treating
 or preventing streptococcal infections such as otitis media,
 meningitis, and bacteraemia
 Example 1; Page -: 113pp; English.
 XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ

Sequence 1152 AA;

Query Match 77.4%; Score 3402.5; DB 23; Length 1152;

Best Local Similarity 79.6%; Pred. No. 1.7e-200;

Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKDNRRVSVYDGSQS--SQKSENL-----TPDQVSQKEGICQAEQI-VIKIT 55
DB 355 SSTAASDNTQSVAKGTSKRPANKSENLSLLKELYDSPAQRYSBDSGLVDFPAKIIISRT 414
QY 56 DQGYVTSRHHYHYNGVVPDPALES-BELIMKDPNYOLKADIVNEVGGYIIKVDGKY 114
DB 415 PNGVAIPGHGHHYF---IPYKSLSALEKTIAR-----MVPISGT- 450
QY 115 YVYLKDAHADNVRTKDEINRQKHQVXDNEKYNVNAVARSQGRYIT-----NDGYV 167
DB 451 -----GSTVST-----NAKPNVYVSLGSLSNPSSLITSKELSSASDGYI 491
QY 168 FNPADIIEDTGNAYIVPHGGHYHYTPKSDLSASELAATAKAGLAGKNMOPQSLSYS-STPS 226
DB 492 FNPKDIIVETATAYIVRHGDFHYIPKSNQIG-----QPTLPNNSLATPS 536
QY 227 PSLPNTGTSHEKHEEDYGFDPANRIIAEDSGFVMSHGDNHNFYFKDLTEEQIKAQK 286
DB 537 PSLPNTGTSHEKHEEDYGFDPANRIIAEDSGFVMSHGDNHNFYFKDLTEEQIKAQK 596
QY 287 HLEEVKTSRSHGLDSSSHQDYPSPNAKEMKDLKKIEKIAKIMKQYGVKRESIVWNKEK 346
DB 597 HLEEVKTSRSHGLDSSSHQDYPSPNAKEMKDLKKIEKIAKIMKQYGVKRESIVWNKEK 656
QY 347 NAIITYPHGDHHDADIDHKPVGIGSHSNYELFKPESGVAKKGNKYITGEELTNVYNL 406
DB 657 NAIITYPHGDHHDADIDHKPVGIGSHSNYELFKPESGVAKKGNKYITGEELTNVYNL 716
QY 407 LKNSFNQNFNLANGKRVSFSPPELEKLGINMLVKLITPGKVLKESGKVGEGV 466
DB 717 LKNSFNQNFNLANGKRVSFSPPELEKLGINMLVKLITPGKVLKESGKVGEGV 776
QY 467 GNIAFELDQPLPQQTQFKYIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDT 526
DB 777 GNIAFELDQPLPQQTQFKYIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDT 836
QY 527 YLRVNPQFAVPKGTDALYRVDFEFGNAYLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPV 586
DB 837 YLRVNPQFAVPKGTDALYRVDFEFGNAYLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPV 896
QY 587 TMANAYLDNOSTYIIVEPILKEKNOTDKPSTLPQFKNKQAQENSKLDEKVEEPTSEKV 646
DB 897 TMANAYLDNOSTYIIVEPILKEKNOTDKPSTLPQFKNKQAQENSKLDEKVEEPTSEKV 956
QY 647 EKEKLETSNNTSNTLEVPVTPVQEKVAFBSYGMKLENVLFNMDGTIELYLPDGE 706
DB 957 EKEKLETSNNTSNTLEVPVTPVQEKVAFBSYGMKLENVLFNMDGTIELYLPDGE 1016
QY 707 VIKKNMADFTGAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPKPVKPPNS 766
DB 1017 VIKKNMADFTGAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPKPVKPPNS 1076
QY 767 TONGMLNPEGNVSDPMLDPALEAPADVPQEKLEKTASYGLDGLSDVIFNMDGTIELR 826
DB 1077 TONGMLNPEGNVSDPMLDPALEAPADVPQEKLEKTASYGLDGLSDVIFNMDGTIELR 1136
QY 827 LPSGVEIKKNL 837
DB 1137 LPSGVEIKKNL 1147

RESULT 15

AAU84056

ID AAU84056 standard; Peptide: 1238 AA.

XX AAU84056;

XX 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, NEW29.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX pneumonia; streptococcal bacterial infection; mutant; mutin;

XX BVH-11-2.

XX Streptococcus pneumoniae.

XX Synthetic.

XX WO2001198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia

Example 1; Page -; il3pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for S. pneumoniae infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological
sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating S. pneumonia nucleic acid in a sample for
diagnosing streptococcal infections. This sequence represents a chimeric
gene created from fragments and variant fragments of Streptococcus
pneumoniae genes, described in the method of the invention.
Note: This sequence does not appear in the specification but has
been created according to information given in the invention.

Sequence 1238 AA;

Query Match

77.4%; Score 3402.5; DB 23; Length 1238;

Best Local Similarity 79.6%; Pred. No. 1.9e-200;

Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKDNRRVSVYDGSQS--SQKSENL-----TPDQVSQKEGICQAEQI-VIKIT 55

DB 441 SSTAASDNTQSVAKGTSKRPANKSENLSLLKELYDSPAQRYSBDSGLVDFPAKIIISRT 500

QY 56 DOGYVTSHGHDHYHYNGKVPYDALFS-EELLMKDPNYOLKADIVNEVKGGYIIRKVDGKY 114
Db 501 PNGVAIPHGDHYHF-----IPYSKLSALEEKIAR-----MVPISGT- 536
QY 115 YVYLKDAHAADNVRYKDEINROKOEHVNDNEKVNNAVARSQGRYTT-----NDGYV 167
Db 537 -----GSTVST-----NAKNEVYVSLGSLSSNPSSLTTSKELSSASDGYI 577
QY 168 FNPADIETGNAYIVPHGCHYHYIPKSDLSASASELAHAKAHLAGKNMOPSQLSYS-STPS 226
Db 578 FNPKDIVEETATAYIVRHGDHEHYIPKSNQIG-----OPTLPNNSLATPS 622
QY 227 PSLPINPGTSHKHEEDGYGFDANRIIADESGFVWSHGDHNNHYFFKDLTDEEQKAAQK 286
Db 623 PSLPINPGTSHKHEEDGYGFDANRIIADESGFVWSHGDHNNHYFFKDLTDEEQKAAQK 682
QY 287 HLEEVKTSHGLDLSLSSHEQDYPNAKEMKDLCKIEEKIAGIMKOYGVKRESIVVNKEK 346
Db 683 HLEEVKTSHGLDLSLSSHEQDYPNAKEMKDLCKIEEKIAGIMKOYGVKRESIVVNKEK 742
QY 347 NAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNL 406
Db 743 NAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNL 802
QY 407 LKNSTNNQNTLANGOKRVSEFPELEKLGINMLVKLIITPDGKVLKYSKGVFGEV 466
Db 803 LKNSTNNQNTLANGOKRVSEFPELEKLGINMLVKLIITPDGKVLKYSKGVFGEV 862
QY 467 GNIANFELDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDT 526
Db 863 GNIANFELDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDT 922
QY 527 YLRVNPQFAPVPGTDALVRVDFEFGHGNAYLENNYKVGEEKLPDKLNQCTTTRTAGNKIPV 586
Db 923 YLRVNPQFAPVPGTDALVRVDFEFGHGNAYLENNYKVGEEKLPDKLNQCTTTRTAGNKIPV 982
QY 587 TFWANAYLQNSTYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 646
Db 983 TFWANAYLQNSTYIVEVPILKEKNOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 1042
QY 647 EKEKLSSETNSTNSLTLEEVPTVPDQEKVAFABSYGKMLNVLNMDGTIELYLPDGE 706
Db 1043 EKEKLSSETNSTNSLTLEEVPTVPDQEKVAFABSYGKMLNVLNMDGTIELYLPDGE 1102
QY 707 VIKKNMADFTGEAPOGNGENKPSNGKSVGTGVNQPTENKPADSLPEAPNEKPVKPPENS 766
Db 1103 VIKKNMADFTGEAPOGNGENKPSNGKSVGTGVNQPTENKPADSLPEAPNEKPVKPPENS 1162
QY 767 TDNGMLNPEGNVGSDPMLDSALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELR 826
Db 1163 TDNGMLNPEGNVGSDPMLDSALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELR 1222
QY 827 LPSGEVKKNL 837
Db 1223 LPSGEVKKNL 1233

Search completed: May 13, 2003, 13:53:44
Job time : 45.8944 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:27 ; Search time 19.0029 Seconds

(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAVALNQHRSQENKNNRVS.....GTIELRPSGEVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-73.*

1: piri:*

2: piri:*

3: piri:*

4: piri:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4274.5	97.2	1039	2 H95115	conserved hypotet
2	4264.5	97.0	1039	2 D97985	hypothetical prote
3	1015.5	23.1	855	2 D98004	histidine Motif-Co
4	1008.5	22.9	853	2 C97985	hypothetical prote
5	989	22.5	839	2 B95136	conserved domain p
6	987	22.5	839	2 G95115	hypothetical prote
7	981.5	22.3	828	2 E98004	conserved domain p
8	975.5	22.2	802	2 C95136	hypothetical 92.4K
9	855.5	19.5	822	2 T46758	hypothetical prote
10	313.5	7.1	182	2 F97985	hypothetical prote
11	212.5	4.8	961	2 G90053	hypothetical prote
12	210.5	4.8	2004	2 F95133	immunoglobulin Al
13	196.5	4.5	940	2 S19702	fibrinectin-bindin
14	194.5	4.4	1963	2 B97002	IgA-specific metal
15	194.5	4.4	2485	1 H71621	serine/threonine-s
16	191.5	4.4	1939	2 T18372	repeat organellar
17	189	4.3	2269	2 T28677	rhodopy protein -
18	187	4.3	1873	2 T30944	surface protein pr
19	186.5	4.2	1849	2 C41859	IgA-specific metal
20	180	4.1	1231	2 T28681	rib protein, - stre
21	180	4.1	1435	2 C37793	erythrocyte-bindin
22	179	4.1	1650	2 T18444	hypothetical prote
23	177.5	4.0	1271	2 A45555	glutamate rich pro
24	176	4.0	2481	2 D90011	FmtB protein (impo
25	175.5	4.0	940	2 AB1744	internalin protein
26	175.5	4.0	1134	2 A60234	IgA Fc receptor pr
27	175.5	4.0	1164	1 FCSOAG	IgA Fc receptor pr
28	174	4.0	1139	1 E64234	cytadherence-acces
29	173.5	3.9	786	2 T18469	hypothetical prote

30	173	3.9	1185	2 A42404	collagen adhesin -
31	172.5	3.9	3488	2 T34418	hypothetical prote
32	172	3.9	710	2 S30154	low-temperature-in
33	172	3.9	4688	2 F82885	hypothetical prote
34	171.5	3.9	940	2 AD3374	internalin protein
35	171.5	3.9	5005	2 F82884	hypothetical prote
36	171	3.9	1018	2 A32192	fibrinectin-bindin
37	170.5	3.9	843	2 H90514	lipoprotein (impor
38	168	3.8	1702	2 A41859	IgA-specific metal
39	167	3.8	1305	2 H41662	150K mating aggreg
40	166	3.8	2401	2 T28676	rhodopy protein -
41	165.5	3.8	347	2 T18674	hypothetical prote
42	165	3.8	1341	2 A37023	IgA-specific metal
43	164.5	3.7	1315	2 T28679	fibrinogen-binding
44	164	3.7	1545	2 B41859	IgA-specific metal
45	163.5	3.7	1694	2 H64106	IgA-specific metal

ALIGNMENTS

RESULT 1

H95115
Conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tetreltin, H.J.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1004

Query Match 97.2%; Score 4274.5; DB 2; Length 1039;
Best Local Similarity 82.3%; Pred No. 4.2e-209;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY	1	CAVALNQHRSQENKNNRVS	YVDGSSSQKSENLT	PDQVVSQKEGIAEQI	VIKIKITDQGYV	60
DB	21	CAVALNQHRSQENKNNRVS	YVDGSSSQKSENLT	PDQVVSQKEGIAEQI	VIKIKITDQGYV	80
QY	61	TSFGDHYHYNGVPYDALF	SELLMKDPNYQLKDADI	NEVKGGYIIKVDGKYVYLKD	120	
DB	81	TSFGDHYHYNGVPYDALF	SELLMKDPNYQLKDADI	NEVKGGYIIKVDGKYVYLKD	140	
QY	121	AAHADNVRTKDEINRQKH	VDNEKVNVAVARSQGR	YTTNDGYVFNPA	DIETGNA	180
DB	141	AAHADNVRTKDEINRQKH	VDNEKVNVAVARSQGR	YTTNDGYVFNPA	DIETGNA	200
QY	181	YIVPHGHHYHYIPKSDLS	ASALAAKAHLAKGNQPS	QLSYSTASDNN	TQSVAKGSTSK	260
DB	201	YIVPHGHHYHYIPKSDLS	ASALAAKAHLAKGNQPS	QLSYSTASDNN	TQSVAKGSTSK	260
QY	224	-----	-----	-----	-----	223
DB	261	PANKSENLOSLKELYDPS	AQRYSESDGLVDFPAKI	ISRTPNGVAIPHGD	HYHFIPYSK	320
QY	224	-----	-----	-----	-----	223
DB	321	LSALEEKIARMPISCTG	STVETNAKPNVWSLGS	LSNPSSLATSKELSS	ASDGYIFN	380
QY	224	-----	-----	-----	-----TPSPSLPINPGT	SHEKHEED 243

```

Db 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 440
Qy 244 GYGFANDRIIAEDESGFVNSHGDNHFFKXDLTTEEQIKAAQKHLEEVKTSNGLDLSLS 303
Db 441 GYGFANDRIIAEDESGFVNSHGDNHFFKXDLTTEEQIKAAQKHLEEVKTSNGLDLSLS 500
Qy 304 HQDYPSNAKEMKDLKKIEKIAKIMQYGVKRESIVVKNKNAIYPHGDHHAADPID 363
Db 501 HQDYPSNAKEMKDLKKIEKIAKIMQYGVKRESIVVKNKNAIYPHGDHHAADPID 560
Qy 364 EHKPVGIGHSNHYELPPEGVAKKEGKNVYTGELTNVNNLLKNSTFNQNTFLANGQ 423
Db 561 EHKPVGIGHSNHYELPPEGVAKKEGKNVYTGELTNVNNLLKNSTFNQNTFLANGQ 620
Qy 424 KRVSFPPELEKKGILNMLVKLITPDGKVLKSGKVGEGVGNIANFELDQYLPQGT 483
Db 621 KRVSFPPELEKKGILNMLVKLITPDGKVLKSGKVGEGVGNIANFELDQYLPQGT 680
Qy 484 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 543
Db 681 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 740
Qy 544 VRVDFEFGNAYLENNYKVGEEKLPKLNQOTTRTAGNKIPVTFMANAYLDNQSTIVE 603
Db 741 VRVDFEFGNAYLENNYKVGEEKLPKLNQOTTRTAGNKIPVTFMANAYLDNQSTIVE 800
Qy 604 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 663
Db 801 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 860
Qy 664 EEPVTPDQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 723
Db 861 EEPVTPDQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 920
Qy 724 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPM 783
Db 921 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPM 980
Qy 784 LDSALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNL 837
Db 981 LDSALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNL 1034

RESULT 2
D97985
hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: D97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174
A:Genetics:
A:Gene: phtE

Query Match 97.0%; Score 4264.5; DB 2; Length 1039;
Best Local Similarity 82.1%; Pred. No. 1.3e-208;
Matches 832; Conservative 2; Mismatches 3; Indels 177; Gaps 1;

Qy 1 CAYALNQHSQENKNNRYSYVDGSSQSKSENLPDQVQSKEGIAEQAEIVIKITDQGVV 60
Db 21 CAYALNQHSQENKNNRYSYVDGSSQSKSENLPDQVQSKEGIAEQAEIVIKITDQGVV 80
Qy 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDAIVNEKGGYIIVKDGKYYVYLKD 120

```

```

Db 81 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDAIVNEKGGYIIVKDGKYYVYLKD 140
Qy 121 AAHADNVRTKDEINRQKHQHVNDKNEKVNNAVARSQGRYTTNDGCVFNPADIIEDTGN 180
Db 141 AAHADNVRTKDEINRQKHQHVNDKNEKVNNAVARSQGRYTTNDGCVFNPADIIEDTGN 200
Qy 181 YIVPHGHGHYIPKSDLSASELAAAHLAGKNMQPSQLSYSS----- 223
Db 201 YIVPHGHGHYIPKSDLSASELAAAHLAGKNMQPSQLSYSSASDNNNTQSVAKGSTSK 260
Qy 224 ----- 223
Db 261 PANKSENQLSKELYDSPAQRYSSESDGLVFDPAKIIISRTPNGVAIPHGDHYHFIYSK 320
Qy 224 ----- 223
Db 321 LSALEEKIARVPISGTGSTVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGYIFN 380
Qy 224 -----TPSPSLPINPGTSHKHEED 243
Db 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 440
Qy 244 GYGFANDRIIAEDESGFVNSHGDNHFFKXDLTTEEQIKAAQKHLEEVKTSNGLDLSLS 303
Db 441 GYGFANDRIIAEDESGFVNSHGDNHFFKXDLTTEEQIKAAQKHLEEVKTSNGLDLSLS 500
Qy 304 HQDYPSNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKNAIYPHGDHHAADPID 363
Db 501 HQDYPSNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKNAIYPHGDHHAADPID 560
Qy 364 EHKPVGIGHSNHYELPPEGVAKKEGKNVYTGELTNVNNLLKNSTFNQNTFLANGQ 423
Db 561 EHKPVGIGHSNHYELPPEGVAKKEGKNVYTGELTNVNNLLKNSTFNQNTFLANGQ 520
Qy 424 KRVSFPPELEKKGILNMLVKLITPDGKVLKSGKVGEGVGNIANFELDQYLPQGT 483
Db 621 KRVSFPPELEKKGILNMLVKLITPDGKVLKSGKVGEGVGNIANFELDQYLPQGT 680
Qy 484 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 543
Db 681 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 740
Qy 544 VRVDFEFGNAYLENNYKVGEEKLPKLNQOTTRTAGNKIPVTFMANAYLDNQSTIVE 603
Db 741 VRVDFEFGNAYLENNYKVGEEKLPKLNQOTTRTAGNKIPVTFMANAYLDNQSTIVE 800
Qy 604 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 663
Db 801 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 860
Qy 664 EEPVTPDQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 723
Db 861 EEPVTPDQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKNMADFTGEAPQGN 920
Qy 724 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPM 783
Db 921 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPM 980
Qy 784 LDSALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNL 837
Db 981 LDSALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNL 1034

RESULT 3
D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Winkler, M.E.

```

J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:CROSS-references: GB:AE007317; PIDN:AAK99864.1; PID:gl5458682; GSPDB:GN00174
C:Genetics:
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.1%; Score 1015.5; DB 2; Length 855;
Best Local Similarity 30.3%; Pred. No. 6.5e-44;
Matches 281; Conservative 136; Mismatches 233; Indels 277; Gaps 31;

QY 1 CAYALNQRSQENKNNRVSYVDGSSQSKSENLPDQVSQKEGIAQEQIVIKITDQYV 60
Db 31 CAYELGLHOAQVTKNNRVSYIDGQATQKTENLPDEVSKREGNAQIVIKITDQYV 90
QY 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIIKVDGKYYVYLKD 120
Db 91 TSHGDHYHYNGKVPYDAIISELLMKDPNYOLKDEDIISEIKGGYIVKVDGKYYVYLKD 150
QY 121 AAHADNVRTKDEINRQKHVVDNE-----KVNNSVAVARSQRYTNDGYVFNPAIDIED 176
Db 151 AAHADNVRTKEINRQKHVVDNE-----KVNNSVAVARSQRYTNDGYVFNPAIDIED 210
QY 177 TGNAYIVPHGGHYHYIPKSDLSASELAQAHLAKGNQPSOLSYSSTPS-----P 227
Db 211 TGDAYIVPHGGHYHYIPKSELASELAQAHLAKGNQPSOLSYSSTPS-----P 270
QY 228 SLIPNPGTSHXHEEDGYGFANDRIIAEDSGFVMSHGDHNNHYFFKDLTEREQIAAKQH 287
Db 271 SVS-NGPTTN----- 279
QY 288 LEEVKTSHNGLDSSHEQDYPNAKEMKDLKKTIEKTAGIMKOYGVKRESI-----V 341
Db 280 ---TNTSN-----SNTSQASQSNIDSLKQLYKLPDSRHVESDGLIFDPAQI 327
QY 342 VRKEKNAIYTPHGHADPIDEHKPVGIGHSNVELFKPEGVAKKEGKNVYVGEILT 401
Db 328 TSTANGVAVPHGDHYHYIPYLSPL-----EKLK 359
QY 402 NVNLLKSTFNQNTFLANGKRVSFSPPE-----LEKKLGINWL 443
Db 360 RIPL-----RYRSNHWVDSRPEQSPQSTPEPSPQAPNPQAPNPIDEKL-VKEA 414
QY 444 VKLITPDGKVLKGVSKVFGEGVGNIANPELDQYLPQOTKYTIASKDYPEVSYDGTFT 503
Db 415 VRKV-CDGYVFE-----NGVPR-----YIPAKDLSAETA 451
QY 504 VPTSLAYKASOTIFYFFHAGDTYLRVNPQFAPVPGTDALRVFDEFGHAYLENNYKV 563
Db 452 KOESLSHKLGAKTDLDPSSDREFY-----NKAYDLARIHQDLDLNNKQGVDFEA 501
QY 564 EIKLPIPKNOGTRTAGNKIPVTFMANAYLDNQSYIIVEVILEKENOTDKPSILPOFK 623
Db 502 -----LDNLERLKDVSSDKVLVDILAFL-----APIRPE-RLGNPAQIYIT 546
QY 624 RNKAQENSL-----DEKVEEPK-TSEK-----VEREKLSETGNSTSN 660
Db 547 DBEIQV-AKLAGKYTDEGYIFDRDITDSEGDYVTPHMTSHWIKKDLSEAEARAAQ 605
QY 661 STLEEYTPDVPQE-----KVAKAESEY-----GHKLENLVNMDGTIELYPSGE 706
Db 606 AVAKEGLTPPSTDHODSGNTAKGAALYNRVKAARKVPLDRMPYNLYQTYVE--VKRGS 663
QY 707 VIKKNMADFTG-----EAPQG-----NGENKPS- 730
Db 664 LIIPHVDHYHNKFEWDEGLVEAPKGSLEDLLATVYVVEHPNPHSDNGFGNASDH 723

QY 731 -----NGKSVGTGVNTOENKPADSLPEAPNEKVPKPENSTDNGMLNPGNVCSDPMLD 785
Db 724 VQRKNQADINQTE-KPNEKPPOTEPETPREKQSE-----KPESPXTEEPPEE 776
QY 786 SALEEAP-----AVDPVOEKLK 803
Db 777 ESPEESPESEPOVETEKVKLE 802

RESULT 4
C97985
Hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
C:Authors: J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:CROSS-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
C:Gene: phtD
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.9%; Score 1008.5; DB 2; Length 853;
Best Local Similarity 31.9%; Pred. No. 1.5e-43;
Matches 295; Conservative 113; Mismatches 263; Indels 255; Gaps 31;

QY 1 CAYALNQRS-QENKNNRVSYVDGSSQSKSENLPDQVSQKEGIAQEQIVIKITDQY 59
Db 20 CSYELGRHQAGQVKKESNRVSYIDGQAGAKAENLPDEVSKREGINAEQIVIKITDQY 79
QY 60 VTSHGHDHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGYYIIKVDGKYYVYLK 119
Db 80 VTSHGHDHYHYNGKVPYDAIISELLMKDPNYQLKSDIVNEIKGGYIVKVDGKYYVYLK 139
QY 120 AAHADNVRTKDEINRQKHVVD-----NEKVNNSVAVARSQRYTNDGYVFNPAIDIEDTG 178
Db 140 AAHADNVRTKEEKRQKHVVD-----NEKVNNSVAVARSQRYTNDGYVFNPAIDIEDTG 199
QY 179 NAYIVPHGGHYHYIPKSDLSASELAQAHLAKG-----NMKPSQ-----LSYSS 223
Db 200 DAYIVPHGDHYHYIPKSDLSASELAQAHLAKG-----NMKPSQ-----LSYSS 259
QY 224 TSPSLPINPGTS-----HEKH-EEDGYGFANDRIIAEDSGFVMSHGDHNN 269
Db 260 TVTPTYHONGENTSSLLRELYAKPLSERHVESDGLIFDPAQITSRANGVAVPHGDHYH 319
QY 270 YFFKDKLPEEQIKAAQKHLSEVKTSHNGLDL-----SSHEQDYPSPN 311
Db 320 FIPYLSQLSPLEKLARIIPLYRSNHWVDSRPEQSPQSTPEPSPQAPNPQAPSPN 379
QY 312 AKEMKDLKKTIEEKTAG--IMKQGVKRESIVNKEKNALIIYPHGDHHDHADPIDEHKPVG 369
Db 380 PTDEK-LYKEAVRKVGDGYVFEENGVP-----YIPAKDLSAETAAG 420
QY 370 IGHSHSNVELFKPEGVAKKE-----GNKV-----TGEELTNV 403
Db 421 IDSKLIAQESLSHLKAGAKKTOLPSSDREFYNKADLLARIHQDLDLNNKQGVDFEALDNL 480
QY 404 VNLKSNFTFNQNTFLANGKRVSFSPPELEKLGINMLVKLITPDGKYLEKYSKVFG 463
Db 481 LERLKDVSDDKVLV---DOLAFAPIRHPERLGRPNQAQITTYDDEIQVAKLAGYTT 536
QY 464 E-----GVCNITANFELDQYLPQOTFKYTIASKDYPEVSYDGTFTVTSIAYKKNASQTIF 518

Db 537 EDGYIFDPDRITSDGDAYVTPHMTSHWIKKDSLSE-----KTSNGLDSL-----SSHEQDYPSPN 311
 Qy 519 YPFHAGTILRVNPOFAVPGKGTDALVRVDFEFGNAYLNNYKVGKIKLPKLNQGTTR 578
 Db 582 YAKEKGLT-----PPSTDHDSGNGTEAKGAEIYNRVKAAK-KVPL-----YIPAKDLSAATAAG 414
 Qy 579 TAGNKIPVTFMANAYLDNOSTYIVPEILEKENQTDKPSILPQFRNKAQENSKLDEKV- 637
 Db 622 ---DRMPYNLQ-----YTVVEV-----GNKXV-----TCEELTNV 403
 Qy 638 BEPKTSEKVEKELSTGNTSNTSLEEVPTVDPVQEVAKFAESYGMKLENLFNMDGT 697
 Db 661 EAPK-----GYSLEDLL-----AT 674
 Qy 698 IEYLPSEGEVIRKKNMADFTGEAPQ-----GNGEN-KPSENGKSVTGTVENOPTENKPAD 750
 Db 675 VKYV-----EHPNERPHSDNGFGNASDHVQRNKNQOATNTE-KPNEEKQOT 722
 Qy 751 SUPEAPNE-----KPVKPNSTNDGMLNPEGVNGSDPMLDSALEEAPAVDPVQEKLEKFT 805
 Db 723 EXPEEDKEHDEVSEPHPSD-----EXENHVGLNPSADNLYKPSSTDTEETEEAE--- 773
 Qy 806 ASYGLGLDSVIFNMGTIELRLPSGE 831
 Db 774 -----DITDEAEIPOVE 785

RESULT 5
 B95136
 Conserved domain protein Sp1174 [Imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: B95136
 R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hoon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, on, J.D.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: B95136
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: GB:AE005672; PID:AAK75283.1; PID:g14972654; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 C:Gene: Sp1174
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.5%; Score 989; DB 2; Length 819;
 Best Local Similarity 31.9%; Pred. No. 1.3e-42;
 Matches 298; Conservative 118; Mismatches 265; Indels 252; Gaps 36;

Qy 1 CAYALNQRS-QENKNNRVSYVDSQSKSENTLPDOVSQKEIQAPQIVIKITDQGY 59
 Db 20 CSVELGRYAGQDKESNRVAYIDGDAQKAENLTPDEVSKREGINAEQIVIKITDQGY 79
 Qy 60 VTSHGHHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYIIVKDGKYYVYLK 119
 Db 80 VTSHGHHYHYNGKVPYDAIIEELMKDPNQLKDSDIVNEIKGYVYKNGKYYVYLK 139
 Qy 120 DAAHADNRTKDEINRQKEHVKD-NEKYNVNAVARSQGRVTTNDGVYFNPADIETDG 178
 Db 140 DAAHADNRTKEIKRQKQERSHNHNSRADNAVAAARAGRYTTDDGYTFNASDIETDG 199
 Qy 179 NAYIVPHGHYHYVYIPKSDLSASLAALKHLAG--NMQPSOLSYSTPSP-----SL 229
 Db 200 DAIIVPHGHYHYVYIPKNSLSASLAALAEAYWNGKQGRSPSSSSSYNANPAQPLSENHL 259
 Qy 230 PINPGTSH-----EKH-EDGYGFDANRIIADSDSGFVMSHGDN 268
 Db 260 TVTP-TYHQNGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGNHY 318

Qy 269 HYFFKKDLTEEQIKAAQKHLEEV-----KTSNGLDSL-----SSHEQDYPSPN 311
 Db 319 HF-----IPYEQNSELEKRIARIIPLYRSNHWVPDSPEEPSQPTPEPSPQAPSN 373
 Qy 312 AKEMKDLKKIEEKIAG--IMQYGVKRESIVVKNENAIIPHGDDHHADPDIHKPVG 369
 Db 374 PIDEK-LVKEAVRKVGQGYFEENGVS-----YIPAKDLSAATAAG 414
 Qy 370 IGHSNVELFKPEEGVAKKE-----GNKXV-----TCEELTNV 403
 Db 415 IDSLAKQESLSHLKTKYDLPSSDREFYNKAYDLLARIHQDLDLNNKQVDFEALDNL 474
 Qy 404 VNLKNTFNQNTLANGQKRVSFSPPELEKKLGIMLVKLITPDGKVLKSVKQVFG 463
 Db 475 LERLKDVSSDKVKLV-----EDILAFAPISHPERLGPNAQIITYTDEIQVAKLAGYTT 530
 Qy 464 E-----GVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMAQOTIF 518
 Db 531 EDGYIFDPDRITSDGDAYVTPHMTSHWIKKDSLSE-----AERAAQA--- 575
 Qy 519 YPFHAGTILRVNPOFAVPGKGTDALVRVDFEFGNAYLNNYKVGKIKLPKLNQGTTR 578
 Db 576 YAKEKGLT-----PPSTDHDSGNGTEAKGAEIYNRVKAAK-KVPL----- 615
 Qy 579 TAGNKIPVTFMANAYLDNOSTYIVPEILEKENQTDKPSILPQFRNKAQENSKLDEKV- 637
 Db 616 ---DRMPYNLQ-----YTVVEV-----KNSLLIPHYDHYHNKIFEFDEGLY 654
 Qy 638 BEPKTSEKVEKELSTGNTSNTSLEEVPTVDPVQEVAKFAESYGMKLENLFNMDGT 697
 Db 655 EAPK-----GYTLEDDL-----AT 668
 Qy 698 IELYL-----PSGEVIRKKNMADFTGEAPQ-----NGENKPSNGKSVTGTVENOPTENK 747
 Db 669 VKYVVEHPNERPHSDNGFGNASDHVQRNKNQOATNTERPSEKPEKTEPEETPREEK 728
 Qy 748 PADSLPAPNEKPY-KPENSTNDGMLNPEGVNGSDPMLDSALEEAPAVDPVQEKLEKFTA 806
 Db 729 PQSEKPEP-KPTEEPPEEP-----SEPOVET-----EKVEEKLEAEED 769
 Qy 807 SYGLGLDSVI-FNMDGTIELRLPSGEVIRKKNLL 838
 Db 770 LLGKIQDPIKSNKAKETL-----TG--LKNLL 795

RESULT 6
 C95115
 Conserved hypothetical protein Sp1003 [Imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hoon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, on, J.D.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: G95115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: GB:AE005672; PID:AAK75120.1; PID:g14972476; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 C:Gene: Sp1003
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.5%; Score 987; DB 2; Length 839;
 Best Local Similarity 32.0%; Pred. No. 1.8e-42;
 Matches 299; Conservative 106; Mismatches 278; Indels 252; Gaps 33;

RESULT 8
C95136
conserved domain protein Sp1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
R:RefSeq: C95136
R:RefSeq: C95136
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holt, I.E.
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <KUR>
A:Cross-references: GB:AF005672; PIDN:AAK75284.1; PID:gl4972655; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.2%; Score 975.5; DB 2; Length 802;
Best Local Similarity 33.6%; Pred. No. 6.3e-42;
Matches 271; Conservative 102; Mismatches 200; Indels 233; Gaps 28;

Qy 1 CAVALNQHRSQENKNNRVSYVDGSSOSKSENLTDPQVSQKEGIAEQIVKIDQGYV 60
Db 6 CSVELGLYQARTYKNNRVSYIDCKQATKENTLTDEYKREGINAQEQIVKIDQGYV 65
Qy 61 TSHGDHYHYNGKVPYDALFSEELLAKDPYQKLDADIVNEKGGYIIKVDGKYVYLKD 120
Db 66 TSHGDHYHYNGKVPYDAIIEELLAKDPYKLDDEIVNEKGGYIKVDGKYVYLKD 125
Qy 121 AAHADNVRTKDEINRQKQEHVKDNE---KVSNNVAVARSQRYTTNDGYVFNPAIDIED 176
Db 126 AAHADNVRTKEETNRQKQESQHREGTTPNDGAVALARSQRYTTDDGYIIFNASDIIED 185
Qy 177 TGNAYIVPHGGHYHYIPKSDLSASELAHAKHLAEVKTSHNGLDLSLSSHEQDYP-----P 227
Db 186 TGDYIVPHGGHYHYIPKNLSASELAHAEFLSGRGNLSNRYIRQSDNTSRTNWVP 245
Qy 228 SLTPNPQTSH-----EKH-EEQYGGFDANRIIA 254
Db 246 SVS-NPQTTNTSNNSTNSQASQNDIDSLKLYKLPQSQRHVESDGLVDPDAQITS 304
Qy 255 EDESGFVMSHGDNHYFFKDLTEEQIKAAQKHLAEVKTSHNGLDLSLSSHEQDYP----- 309
Db 305 RTARGVAVPHGDHYHYIPYSQMSSELEERIIIPLYRSNHWVDPDRPEQSPQPTPEPS 364
Qy 310 -----SNAKEMKDLKXKIEKIAQIMKQYGVK- -----ESIVYNNKKNAI 349
Db 365 PGPOPAPNLKIDSNSLVSQLVRKGE--GVYFEKGISRVYFAKDLPSSTVKNLSKL- 421
Qy 350 IYPHGDDHHDADIDEHPKVPVIGSHSNYELFKPEEGVAKKGGKVVYTGELTNVLL-- 407
Db 422 -----SKOESVSHLT-----AKKENVAPRDQEFYQKAYNLLTE 455
Qy 408 -----KNSTF-----NNQNFLLANGKRV-----SFSFPELEKKLGINMLVKL 446
Db 456 AHKALFENKGRNSDFQALDKLLERLNDESINKELVDLLAFAPITHPELG----- 508
Qy 447 ITPDGKLVLEKVSQVFGEGVGNANFELDQYLPQGTFFKTYIAS-----KDYPEVSDGT 502
Db 509 -KPNSQL-----EYDEVRVIAQADKYTTSDGYIFDEHDIISDEGDA 550
Qy 503 TVPISLAY-----KVASOTIFYPFHAGDYLAVNQAFAVPGKTDALVRVFE 549
Db 551 YVTPHMGHSHWIGKDSLDKQVAAQA--YTKKG-----ILPPSPDADKANPT 598

Qy 550 FHGNAYLENNYKVGKILPIPKLNQGTTRTA---GNKI-----PVTFMANAYLDNQ 597
Db 599 GDSAAAIYNRVK-GEKRIPLRVLRYMVEHTVEVXNGNLIIPHKDHYHNKF---ANFDDH 654
Qy 598 S-----TYIVEVP-----ILEKENOTDKPSILPOPKRN 625
Db 655 TYKAPNGYTTLEDLFAIKYVYVHEPDERPHSNDGMGNASEHVLGKKDHSEDPN-----KNF 709
Qy 626 KAQNSKLDKVEPEPKSEKVEKEL 651
Db 710 KADE-EPVEETPAPEVPQ-VETEKV 733

RESULT 9
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heynemann, J.; Schnitzler, N.; L
Infect. Immun. 67, 871-878, 1999
A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attach
A:Reference number: 224091; MUID:99115568; PMID:9916102
A:Accession: T46758
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPE>
A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624
A:Experimental source: strain R268
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 19.5%; Score 855.5; DB 2; Length 822;
Best Local Similarity 29.0%; Pred. No. 8e-36;
Matches 270; Conservative 143; Mismatches 277; Indels 241; Gaps 45;

Qy 2 AYALNQHRSQENKNNRVSYVDGSSOSKS---ENLTDPQVSQKEGIAEQIVKIDQGY 59
Db 22 SVLOKQHHMGLATKDNQIAYIDDSKGYKAPKTKNTMDQISAEESIAEQIVKIDQGY 81
Qy 60 VTSHGDHYHYNGKVPYDALFSEELLAKDPYQKLDADIVNEKGGYIIKVDGKYVYLK 119
Db 82 VTSHGDHYHYNGKVPYDAIIEELLMTDPNHFQSDVINEILDGYVIKNGYVYLK 141
Qy 120 AAHADNVRTKDEINRQKQEHVKD-NEKVSNNVA-----VARSQRYTTNDGYV 167
Db 142 POSKKNRTKQIAEQVAKTKERKGLAQVAHLSKEEYAAVNAEKROGYTTDDGYI 201
Qy 168 FNPADIEDTGNAYIVPHGGHYHYIPKSDLSASELAHAKHLA---GNMOPQSLSYST 224
Db 202 FSPTDIIDGLDAYLVPHGNHYHYIPKDLSPSELAAQAYWSQKRGARPS--DYRPT 259
Qy 225 PSP---SLPI-----NPGTSHE-----KH----- 240
Db 260 PAPGRKAPIDVTNPGQGHQPDNGYHPAPPENDASQNKHQHDEFKGTFKELLDQL 319
Qy 241 -----EEDGYGFANRIIAEDSGFVMSHGDNHYFFKDLTEEQIKAAQKHLAE 290
Db 320 HRDLKRYHVEEDGLTFEPTQVKSNAFYVPHGDHYHYIPRSQLSPLEMELADRYLAG 379
Qy 291 VYTSNGLDLSLSSHEQDYPNAK-----EMKDLKIEKIAQIMKQYGVKRESIVV 342
Db 380 -QTDON--DSGSDHSK--PSDKVTHFLGHRKAYKGLDGKPYDTSDAIVFSKESI-H 433
Qy 343 NKEKNAIYPHGDHHDADIDEHPKVPVIGSHSNYELFKPEEGVAKKGGKVVYTGELTN 402
Db 434 SVDKSGVFAKHGCHYH-----IGFGL-EEQELDEVANWV-KAQG---ADELVA 479
Qy 403 VYNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGITNMLVKLITPDGKV--LEKVSQK 460
Db 480 ALD-----QEQGKEKPLFD-----TKKVS-----RKTKDGKGVYINPKDGK 516
Qy 461 VFGEGVGNIANFELDQYLPQGTFFKTYIASKDYPEVSDGTFTVPTSLAYKMASOTIFYP 520

Db 517 DY-----FYARYQLDLTQI--AFAEQELMLKDKKHRYD---IVDTGIEPRLAVDVSSLP 566
Qy 521 FHAGDTYLRVNPQFAVPGKTDALVRVFEDEHGNAY--LENNYKGEIKLIPKLNQGTTR 578
Db 567 MHAGNATYDTGSSVIPH-----IDHIHVYPYSLTRN-QIATIK----- 605
Qy 579 TAGNKKIPVTFMANAYLNDQSTYIVVEP-----ILEKNQTDKPSILQFQKRAQNSKL 633
Db 606 -----YNNQHPREVDPVWMSKPGHEESGVPINV-----TPL 636
Qy 634 DEKVEEPK-----TSEKVEKEKLETSNSTNSTLEEVPTVDPQEKVAKFAESYGMKLE 688
Db 637 DKRAGMPNWIHSAEEVOK-ALAEGRFAADPGYI-----FDP-RDVLAQ-EFVWK-- 685
Qy 689 NYLFNMDGTIELYLPSEGVIKK-NMADFTGEAPQNGENKPSKGVSTGTGVNQPTENK 747
Db 686 -----DGSFIPRADGSSLRITNKSQAEWQAOQELLAKKNAGDATDT--DKPEEK 737
Qy 748 PADSLPEAPNEKP---VKPENSTNGMLNPEGVSDPMLDSALEAPAVDPVQEKLEKF 804
Db 738 QADKSNE--NOQGEASKEEKESDDFI-----DSLPGYGLDRATLEHDHINQLAQK- 785
Qy 805 TASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 786 -----ANIDPKYLIFQPEG-VQFYKNKGCELY 810
RESULT 10
F97985
hypothetical protein phtB-truncation [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:g15458517; GSPDB:GN00174
C:Genetics:
A:Gene: phtB-truncation
Query Match 7.1%, Score 313.5; DB 2; Length 182;
Best Local Similarity 39.4%; Pred. No. 2.6e-09;
Matches 76; Conservative 20; Mismatches 62; Indels 35; Gaps 6;
Qy 153 VARSQGVTTNDGYVFNADIIETDGNAYIVPHGGHYHYIPKSDLSASELAARAHLAGK 212
Db 3 VTPNNGVSAVDGCVFNENDIVDTGDAYIVRHGDHYHYIPKSSL----- 47
Qy 213 NMOPQSLSYSPSPSLPI--NPGTSHEKHEEDGYGFDANRIIAEDESFGVNSHGDHNY 270
Db 48 NNPPSHNSNTEEVGSSSVLSNPRLRVHHEEDGHHGFDANRIISDSGCVIPHGDHNY 107
Qy 271 PFKKDI.TEEQIKAAOKHLEEVKTSNGLDLSSSHQDQVSNKAKMKLDKKTIEKIA--- 327
Db 108 -----IKVQTGYE--AALKNKIPSLQS-----NYPGTFDEKAVLAKVDQLLDSR 152
Qy 328 GIMQYGVKREST 340
Db 153 SIYKDKPIEQRI 165
RESULT 11
G90053
hypothetical protein fnbB [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G90053
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G90053
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-961 <KUR>
A:Cross-references: GB:BA000018; PID:g13702452; PIDN:BA843593.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fnbB
Query Match 4.8%; Score 212.5; DB 2; Length 961;
Best Local Similarity 21.4%; Pred. No. 0.0042;
Matches 176; Conservative 110; Mismatches 287; Indels 251; Gaps 42;
Qy 26 QSSQKSE-NLTPDOVSQKEGIAEQIVIKITDQGYVTSHG--DHYHYNGKVPYDALFSE 82
Db 290 QGSQKVEVTLQCNKVSKEFDI---KYLDGVKDRMGVTNGRIDTLNKEGK-----FSH 340
Qy 83 ELLMKDENYOLKADIVNEVKGGYLIKVDG---KYVYLKDAAHADNVRTKDEINPKOE 139
Db 341 FAYVKNQSLTSVTIGQVTSQTKOSANNPTVKYKHIGSDLAESVAKLD-DTSKFE 399
Qy 140 HVKDNERXVNSNAVARSQGRYTTNDGYVFNPAADIIEDTGNAYIVPHGGHYHYIPKSDLSA 199
Db 400 DV--TEKNLS-----YTSNGGYTLNLGDL--DNSKDYIVKYEGEY-----DQNA 440
Qy 200 SELAAKAHLAKGNMQPSQLSYSTSPSPSPINPGTSHKHEHEEDGYGFD--NRITAEDE 257
Db 441 KDL-NFRTHUSGHYKHYPPYPPYPPVQVTWNGVAFYNNAGKDGCKDPNPIEKSE 499
Qy 258 SGFVMSGHDNHHFFFKKDLTEQIKAAQKH-----LEEVKTS-----HNGDLSLSHE 305
Db 500 -----PIDLDIKSEPPVEKHELTGIEESNDSKPIDFEYHTAVEGAEGA 544
Qy 306 QDYPNNAKMKLDKKTIEKIAGIMQYGVKRESIVVYNNKNAIIPHGDDHHAIDPEH 365
Db 545 E-----GIIE---TEDSIHVDFEEST---HENSXHHADVVEYE 577
Qy 366 KPVGIG---HSHSNYELFKPE--EGV-----AKEGNKVYTGELTNVNL----- 406
Db 578 EDTNPGGGQVTTESNLVFEDEESTKGIIVTGAUSDHITVEDTKEYTTE--SNLIELVDLP 635
Qy 407 -----LKNSTFNQNT---LANGQKRVSFSPFPELEKLGINMLVKLITPDGKVL 454
Db 636 EEHQAGQPIIEETENHHHSHSGLGTENGHNGYGVIDEIEENSHVDIKSEL----- 687
Qy 455 EKVSQGVFEGVGNIAFELDQPLPGQTFKYLTSKDYPEVSYDGTFTVPTSLAYKMAS 514
Db 688 -----GYEGGQNSGN-QSFEED-----TEDKPKYEGGNI----- 717
Qy 515 QTIFFPFAHGDYTLRVNPQFAVPGKTDALVRVFEDEHGNAYLENNYKGEIKLIPKLNQ 574
Db 718 -----VDIDF-----DSVPQIHGQNNQNSFDEETDEEK-----PKYEQ 751
Qy 575 GTTGTAGNKIPVTFMANAYLNDQSTYIVVEPILKEKNQDKPSILQFQKRAQNSKLD 634
Db 752 G-----GNIIDIDSDVPQIHGFNKH---NEIIEEDTNKDKPNY--QF-----GHNNSVDF 797
Qy 635 EKVEEPTSEKVEKEKLETSNSTNSTLEEVPT-----VDPVQEKVAKFAESYGMKLEN 689
Db 798 EEDTLPKVSGQNEGQQTTEEDTTPPTPTPEVPSEPTPTPTPEVPSEGEPTPKPE- 856
Qy 690 VLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGE-----NKPSENGKYSTG 737
Db 857 ----VPSEPTFVPP-----TPVPESEPKVPPPAKEPKKPSKPEVQCGKVTP 901

QY 492 DYPEVSYDGTFTVPTSLAYKMASQITFYFFHAGDTYLRVNPQFAVPRKGTDALRVFDEHF 551
DB 696 DRPKYEOGNI-----VDIDF-----DSVPQIHGQNN 722
QY 552 GNAYLENNYKVGKIPKLPKLNQGTTRAGNKIPVTFMANAYLDNQSYIIVEVPIEREN 611
DB 723 GNQSPEDTEKDK-----PKYEQG-----GNIDIDFDSVPHIHGFKNH---TEIEEDT 769
QY 612 QTDKPS-----ILPQFK-RNKAQENSKLD-----EKVEEPKTS 643
DB 770 NKDKPNYQGGHNSVDFEEDTLPOVSGHNEGQOTIEEDTTPVPTPTPEVPSEPTP 829
QY 644 ERVEKEKLSGTNSNSTLEEVPT-----VDPQVKFAKFAESYGMKLENVFNMDGTI 698
DB 830 TPEPTPEVSEP-ETPTPTPEVPTPEPKPIPPAKEEPK----- 867
QY 699 EYLYPSGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVB-NQ-----PTENKPA-- 749
DB 868 ----PS-----KPVEQGVVTPVIEINEKRAVAVPTKKAQSKK 901
QY 750 DSLPEAPNEKPKVPKPNSTNDGML 772
DB 902 SELPTGGEE-----STNGML 918
RESULT 14
B98002
Iga-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: B98002
R.Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A57872; MUID:21429245; PMID:11544234
A:Accession: B98002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1963 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:g15458662; GSPDB:GN00174
C:Genetics:
A:Gene: iga
A:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 4.4%; Score 194.5; DB 2; Length 1963;
Best Local Similarity 22.0%; Pred. No. 0.098;
Matches 165; Conservative 85; Mismatches 220; Indels 281; Gaps 44;

QY 186 GGHYIYKPSDLSASELAALAKLAKNMQPSQSY-----SSTP 225
DB 47 GVHYKYADSELSSEE-----KKQLYDIPTYVENDDETYYLVLYKLSNQ 92
QY 226 SPSPINPOTSHEKE-EDGYGFANDRIIAEDSGFVMSHGDNHFFPKDLTEBQIAKA 284
DB 93 LAELP-NTGSKNERALYAGASLALGILI-----FAVS-----KKVKNKTV--- 134
QY 285 QHLEEVKTSHG-----LDSLSH-----EQDPSNAKEMKDLKKIE-EKIAGIMKQYG 334
DB 135 -LHLVLVAGMGVLSVHALENHLLNNTDYELTSCEKLPKPKEISGYTVIGYKEGK 193
QY 335 VKRESIVNKKENAI-----IYPH-GDH-----HHADPIDEHKPVGIG----- 371
DB 194 TSDPEVSNQEKSAATPTKQKQVYNVTNPNFVDHSTVQAIQEQTPVSSSTKPTPEVQVVEK 253
QY 372 -----HSHSNYELFKPEEGVAKKGNKVYPT-----GEELTNV 404
DB 254 PFSTELINPRKEKQSSDQEQALAEKHNET-KKEEKISPKETGVNTLNQDVELSQL 312
QY 405 N-----LLKNSTFN-----NQNTLANGQKRVSFSPPELEKKLINM-LVKLIT-- 448

DB 313 NKPELLIREETIETKIDFQBEIQENPOLAEGTVRV-----KQEGKLGKVEIVRIEFSVN 366
QY 449 -----PDGKVLKYSGR--VFGE-----GV-----GNIANFELDQYLP 480
DB 367 KEEVSREIVSTTAPSPRIVEKGTGKTKQVKEQPETGVEKHQVQSGAIVEPAI-QPELP 425
QY 481 GQTKYIATSKDYPEVSDGTFTVPTSLAYKMASQITFYFFHAGDTYLRVNPQ-----FAV 536
DB 426 -----EAVVSDGGEPEVOP-----TLPEAVV-----TDRGET--EYQPESPDPVVS 464
QY 537 PKGTDALVRVDEFGHAYLNNYKVGKIK--LPIPKL-NQGITRTAGNKIPVTETMANAY 593
DB 465 DKGEPEQVAPLPEYKGN-----IEQVXPETPVENTKEQGEKPT--EVPV----- 507
QY 594 LDNSTVIVVEPILEKENQDKPSILPOFKRN-----KAQENSKLDEKVEEPEKTSKV----- 646
DB 508 -----KPTETFPVNPNEGTTGTSI--QEAENPVQPAEESTINSEKVSPTSSENTGEVS 560
QY 647 -----EKEKLSGTNSNSTLEEVPTDPVQEKVAKFAESYGMKLENVLFN 693
DB 561 SNPSDSTISVGESNKPENHNSKNSENSEKTYVEVP-VNP----- 597
QY 694 MDGTIELYPSGEVIKKNMADFTGEAPQNGENKPSNGKV---STGTVENQPTENKP-- 748
DB 598 NEGTV-----GTSNQETEKVPQPAEETQTNISGKIANTENTGEVSNKPSDSRPPV 646
QY 749 ADSLPEAPNEKPKVPKPNSTNDGMLNPE 775
DB 647 EESNQPEKNGTATPENSNGNTTSENGQTEPE 677
RESULT 15
H71621
serine/threonine-specific protein kinase (EC 2.7.1.1) PFB0150c - malaria parasite (P1
C:Species: Plasmodium falciparum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: H71621
R.Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71660; MUID:99021743; PMID:9804551
A:Accession: H71621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2485 <GAR>
A:Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AACT71820.1; PID:g384
A:Experimental source: clone 307
C:Genetics:
A:Gene: PFB0150c
C:Superfamily: malaria parasite serine/threonine-specific protein kinase PFB0150c; pr
C:Keywords: phosphotransferase
F:2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.4%; Score 194.5; DB 1; Length 2485;
Best Local Similarity 20.0%; Pred. No. 0.14;
Matches 173; Conservative 120; Mismatches 316; Indels 265; Gaps 41;

QY 5 LNHRSGENKDN-NRVSIVDPSSQSKSENLTDPQVSQKEGI-----QAEQIVIKIT 55
DB 796 ITPYESDKENKNSNIKSMDFNVVVKRKSNNLYNLSNRDSTVDMHNKYNSEEVINQRT 855
QY 56 DQGVVTSHG--DHYHYNGKVPYDALFSELLMKDPNQLKADIVNEVKGVIIRKVDGK 113
DB 856 NKIYELNKRIRNTKLYSMDEIFKVSLEKKYIDNTISNNMERTYRNEMINEXLSKMDDI 915
QY 114 YVYVLKDAAHADNVRTKDEINRQKQEHVKD-----NEKVSNAVAVASQGRYTTNDGYVF 168
DB 916 LYPCKDKNLSNMSCFVLIENNISRENEKSSVTLNKKKEN--MFCVGRGLCHHGKM- 972
QY 169 NPADLIEDTGN-----AYIVPHGGHYHYIPKSDLSASELAALAKLAKNMQPSQ----- 218

GenCore version, 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 : Search time 9.85337 Seconds
(without alignments)
3535.857 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAVALNQHRSQENKNNRVS.....GTIELRLPSCVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	4.2	1849	1 IGA4_HAEIN	P45386 haemophilus
2	180	4.1	1435	1 EBAL_PLAFC	P19214 plasmodium
3	175.5	4.0	1164	1 BAG_STRAG	P27951 streptococc
4	174	4.0	1139	1 HWI_MYGE	Q49413 mycoplasma
5	173	3.9	1183	1 CNA_STAAU	Q53654 staphylococ
6	172	3.9	710	1 L778_ARATH	Q06738 arabidopsis
7	171	3.9	1018	1 FNBA_STAAU	P14738 staphylococ
8	168	3.8	1702	1 IGA2_HAEIN	P45384 haemophilus
9	165.5	3.8	1637	1 MRSP_STAAU	P80544 staphylococ
10	165	3.8	1541	1 IGA1_HAEIN	P42782 haemophilus
11	164	3.7	1545	1 IGA3_HAEIN	P45385 haemophilus
12	163.5	3.7	1694	1 IGA0_HAEIN	P44969 haemophilus
13	162.5	3.7	715	1 STC2_STAAU	P17855 staphylococ
14	159.5	3.6	2195	1 SC16_YEAST	P48415 saccharomyc
15	158.5	3.6	1701	1 MSPI_PLAFM	P08569 plasmodium
16	157.5	3.6	3924	1 ANK2_HUMAN	Q01484 homo sapien
17	157	3.5	1381	1 YBE7_YEAST	P34216 saccharomyc
18	156	3.5	1453	1 Y373_BOVIN	Q9tu23 bos taurus
19	155.5	3.5	1701	1 NSPI_PLAFC	P13819 plasmodium
20	155	3.5	1073	1 RESA_PLAFC	P13830 plasmodium
21	154	3.5	1750	1 Y832_METJA	Q58242 methanococ
22	154	3.5	1790	1 USO1_YEAST	P25386 saccharomyc
23	153.5	3.5	1577	1 HLYA_PROMI	P16466 proteus mir
24	153.5	3.5	1726	1 MSPI_PLAFC	P04934 plasmodium
25	153.5	3.5	1726	1 MSPI_PLAFC	P50495 plasmodium
26	153.5	3.5	2748	1 NUM1_YEAST	Q00402 saccharomyc
27	153	3.5	1744	1 TANA_XENLA	Q01550 xenopus lae
28	152.5	3.5	1630	1 MSPI_PLAFC	P04932 plasmodium
29	152.5	3.5	1639	1 MSPI_PLAFC	P04933 plasmodium
30	152.5	3.5	1658	1 YM67_YEAST	Q03661 saccharomyc
31	152	3.5	747	1 Y030_UREPA	Q9ptb5 ureaplasma
32	149.5	3.4	857	1 NFM_CHICK	P16053 gallus gall
33	148.5	3.4	1159	1 N124_SCHPO	Q09904 schizosacch

34 148.5 3.4 1287 1 VAC2_HELPY Q48245 helicobacte
35 148.5 3.4 1466 1 SPA2_YEAST P23201 saccharomyc
36 147.5 3.4 1167 1 SCAL_STRPY P15926 streptococ
37 147.5 3.4 1233 1 YE16_YEAST P43597 saccharomyc
38 147.5 3.4 1433 1 REST_CHICK O42184 gallus gall
39 147 3.3 818 1 PTK2_YEAST P47116 saccharomyc
40 146.5 3.3 3135 1 S230_PLAFC Q08372 plasmodium
41 146 3.3 1085 1 IFH1_YEAST P39520 saccharomyc
42 146 3.3 1240 1 YN71_YEAST P53935 saccharomyc
43 146 3.3 1332 1 SPT7_YEAST P35177 saccharomyc
44 146 3.3 1636 1 BUD3_YEAST P25558 saccharomyc
45 145.5 3.3 774 1 GLYE_PLAFC P02895 plasmodium

ALIGNMENTS

RESULT 1
IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTI HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -!- PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-|-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: M87491; AAA24968.1;
DR MEROPS; S06.001;
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AFA CRC64;

```
Query Match          4.2%  Score 186.5;  DB 1;  Length 1849;
Best Local Similarity 18.0%  Pred. No. 0.12;
Matches 166;  Conservative 128;  Mismatches 344;  Indels 283;  Gaps 35;

QY 12 ENKNNRVSYVDGSO--SSOKENLTPDOVSOKEGIOAEQIVIKITDQGYVSHGDHYHY 70
DB 471 KGNKGLLVGDGTIVLKOKADANKKQVAFQVGVISGRSTLVNDKQVDPNSYFGR 530
QY 71 NCKVPYDA-----LFESEELMKDPN-----YOLKADI 98
DB 531 GGRDLNGSLTFDIRNIDDCARVVNHNMTNITITGESLITNPWTITSYILEAQDD 590
QY 99 VNEV-----KGGYIIKVDGKYVYLKDAHADNV-----RTKDEINRQ 136
DB 591 DPLAIRSIPYRQLYFNQDNRSYITLKRGASTRSELPONGSENNWLMYGMTSDEAKRN 650
QY 137 KOEHVKNKNSVAVARSQRYTTNDGYVFNPAIIEDTGNAYIVPHGGHYHYIPKSD 196
DB 651 VNNHI--NNERMNG-----FNGYGEETKATQNGKLVNTENG-----KSD 689
QY 197 LSASELAATAKAHLACKNMOPSSOLSYSSSTPSPSLNPINPQTSHEKHEEDYGFDAHLIAD 256
DB 690 QNRFLITGTLNGDLNVEKGLFSLGRPTPHARDIAGISSTK--KDPHFTENNEVVVED 747
QY 257 E-----SGFVMSHGDHNNHYFPKDLTBEQKAAQKH 287
DB 748 DWINRFKATTWNVTGNASLYSGRNVANTSNITASNAAQVHIGYK---TGDTCVRSYD 804
QY 288 LEEVKTSHNGDLSLSHEODYPSNAKEMKDLKKI-----EKTAGIMKQYGVKRESIIV 342
DB 805 TGVV--TCHNSLSEKALNSFNENLNGVNLTENASFTLGRKANLFGTIGSITSO---V 859
QY 343 NREKNALIPHDGHHHADIDDEKPGVI--GHSNSYELFKPEEGVAKKEGKVTGEEL 400
DB 860 NLKENS-----HWHLTGNSVYNOLNTNGHILN-----AONDANKVITYNTL 902
QY 401 TNNVLLKNTF-----NNQNETLANGQRVSPFPPELEKRLGIMLMLITPDGK 452
DB 903 T-VNSLSGSGSYFYVYDFTNKSNKVVN--KSATGNETLOVADKTGEPHNNELIFDAS 959
QY 453 VLEKVGKVGEGVGNGIANFELDQYLPQOTKYTIASKD-----YPEYSYDGTFTVPT 506
DB 960 NATRNLEV-----TLANGSVDR-----GAWKYLRNVNGRYDLYNPEV-----998
QY 507 SLAYKASOTIFYPHAGDTLYLRNP--QFAPVPGTDALVRVDFEFGHGNAYLENNYKVG 563
DB 999 -----EKRNQTV-----DTNITITPNDIQADAPSAQ-----SNNEETA 1031
QY 564 EIKLPKLNQGTTRTAGNKIPVTFMANAYLONQSTYIIVEVILKENOTDKPSILPQEK 523
DB 1032 RVETPPVPAPAPATESAIASEQPET-----RPAETAQPA-----1064
QY 624 RNKAQENSKLDEKVEEPKTSKYEKEKLSSETGNSSTNSLLEEVPTVDPVQEKVAKFAESY 683
DB 1065 -----METNTANSTETAPKSDTATOTENPNSVPS-----ETTEKVAENP 1106
QY 684 GMKLENVLFNMDGTIELYLPSSGV IKKNMADF-----TGEAQQGNGENKPSNGKVSIGT 738
DB 1107 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSECKTEQTQ-----TAE 1161
QY 739 VENQPT-----ENKPADSLPEAPNEKPV-----KPNSTDNGLMNPENGVSDPM-----783
DB 1162 TKSEPTSVTVSNGQPEKTVSQSTEDRVVVEKEEKAKVETEETQKAPQVTSPEPKQAPZ 1221
QY 784 -----LDSALEEAPAVDPVQ 798
DB 1222 APEEVPDTTNAEEAQAALQOQTQ 1242

RESULT 2
EBAL_PLAFC
ID EBAL_PLAFC
AC P19214;

PRT: 1435 AA.
STANDARD;
```

Db 1289 RYNL-----YDKKLDLDLYENRNDSTTKELIKKLABEINKCENEISVK 1330

RESULT 3

BAG_STRAG STANDARD; PRT; 1164 AA.
 AC 27951;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Iga FC receptor precursor (Beta antigen) (B antigen).
 GN BAG.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
 RC STRAIN=LA239;
 RX MEDLINE=91312121; PubMed=1857207;
 RA Jeriastroem P.G., Chhatwal G.S., Timmis K.N.;
 RT "The Iga-binding beta antigen of the c protein complex of Group B
 streptococci: sequence determination of its gene and detection of two
 binding regions";
 RL Mol. Microbiol. 5:843-849(1991).
 RN [2]
 RP IDENTIFICATION OF IG-LIKE DOMAIN.
 RX MEDLINE=97035265; PubMed=8880921;
 RA Bateman A., Eddy S.R., Chothia C.;
 RT "Members of the immunoglobulin superfamily in bacteria";
 RL Protein Sci. 5:1939-1942(1996).
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (potential).
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X59771; CAA42442.1; .
 DR PIR; S15330; FCSOAG.
 DR InterPro; IPR004829; Csurface_antigen.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003599; Ig.
 DR Pfam; PF00746; Gram_pos_anchor; 2.
 DR ProDom; PD153432; Csurface_antigen; 1.
 DR SMART; SM00409; IG; 1.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR TIGRFAMs; TIGR01168; YSRK_Signal; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
 FT Immunoglobulin domain.
 FT SIGNAL 1..37
 FT CHAIN 38..1135 IGA FC RECEPTOR.
 FT PROPEP 1136..1164 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 434..534 IG-LIKE DOMAIN.
 FT DOMAIN 139..438 IGA-BINDING (POTENTIAL).
 FT DOMAIN 439..826 IGA-BINDING (POTENTIAL).
 FT DOMAIN 827..945 PRO-RICH REPEATS.
 FT SITE 1132..1136 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1135..1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;

Query Match 4.0%; Score 175.5; DB 1; Length 1164;
 Best Local Similarity 18.5%; Pred. No. 0.23;
 Matches 163; Conservative 128; Mismatches 295; Indels 295; Gaps 40;

QY 116 VYLKDAHADNVRTKDEINRQKQEHVKNEKYNVAVARSQGRVTTNDGYVFNPADLIE 175

Db 29 LFMGVAHA-----SELVKDSVKTTEVAAPKPYPSMAQTQGG----- 65
 QY 176 DTGNAYIVPHGGHYHYIPKSDLSASLAAKAHL-----AGKNMQPSQLSYSTPSPSLPI 231
 Db 66 -----NNSSSSELETTKMEIPTTDIKRAVEPVEKTAGTSA----- 101
 QY 232 NPGTSHEKHEEDGYGFDANRIIAEDESQFVAKSHGDHNYFFKKDLTEQIRAAQKHLEEV 291
 Db 102 ---TTGKREKOLQWKN--LKNDVNTILSHEQNE--FKTKIDETNDSALLELEN 154
 QY 292 KTSNGLDLSLSHEQ-DYPSNAKEMKDL---DKTEEKIAGIMKQYGVKRESI----- 340
 Db 155 FNETRLLHIKQHEEVKDKAKQOKTLKQSDTKVD--LSNIDKELNHQKSOVKMAEQK 212
 QY 341 -VYNKEKNAILYPHGD-HHHADPIDHRPVGIGSHSHSYELFKP-----BEGVAKK 389
 Db 213 GTNEDKDSMLKKIEDIRKQAOQADKEDAEVYKREELGLFSSTKAGLDQDEIQEHVAKK 272
 QY 390 EGNKYVTGEELTNVNNLKN-----STFNQ-----NFTLANGOK----- 424
 Db 273 TSSEENTQKVDHYANSLQNAQKSLEELDKATTNEQATQVKNQFLENAQKLKEIQPLIK 332
 QY 425 -----RVSTFPPELEKKLGINM---LVKLITPDGKVLKVSQKV----- 461
 Db 333 ETNVKLYKAMSESLEQVEKELKHSEANLEDLVAKSKSIVREYEGKLNQKNLPKOLE 392
 QY 462 -----FGEVGNI-ANFELDQYLPFGOTFKYTIASKD-----YPE--VSYDG-- 500
 Db 393 EEAHSLKQVVEDFRKFKTSEQVTPKKRVRDLAANNENQOKTELTVSPENITVYEGED 452
 QY 501 -TFTVPTSLAYKMASQTIYFPFHAGDTYLRNPOFAVPKGTDALVRVDFEFHGNAYLENN 559
 Db 453 VKFTVTA-----KSDSKT---TLDFSLLTKYNPSVS-----DRISTNYKTNT--DN 494
 QY 560 KVGSIKLPKIPKNGTDT-----AGNKIPVTFMAYLDNQSTYIVVEPILKEKNT 613
 Db 495 HIAIEITINKLNEQSQTIVLAKDSDNVEKTF-----TTIVOKKEEK- 539
 QY 614 DKPSILPQFKRKAQENSKLDEKV-EETKSEKVEKEKLSQTN-----STNSSTLEEV 666
 Db 540 -----QVPKTPGKDSKTEEPQEPKSNQOLQELIKSAQOELEKLEKAIKELMEQ 592
 QY 667 PTV-----DPVQEKVAKF-----ASYGNKLENVLFNMDGT 697
 Db 593 PEIPSNPYGIQKSWESQKEPIQEAITSFKKIIGDSSSKYTYTHYFNKYKSDFMNYQLH 652
 QY 698 IELYLPSEVGIKKNMADFTGEAPQNGENKPSQKSVT-----GTVENQKTE----- 745
 Db 653 AQM-----EMLTRKYVQYMNKYPDNAETKIFESDMKRTKEDNYGSLNDALKGYFEKYP 707
 QY 746 ---NK-----PADSLPEAPNEKVPKPDENSTDGMLNPEGNVGSDDPMLDSALBEA----- 791
 Db 708 LTPFNKIKQIVDDLDKQVQDQAPAPI-PENSE-----MDQAKERAKIAV 750
 QY 792 -----PAVDPPQEKLEKFTASYGLGL-----DSVIFNMD 820
 Db 751 SKYMSKVLGDGVHQLQKNNSKIIVDLFKLEAIKQOTIFDID 792
 RESULT 4
 ID HMW1_MYCGE STANDARD; PRT; 1139 AA.
 AC Q49413; Q49365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytadherence high molecular weight protein 1 (Cytadherence accessory
 protein 1).
 DE protein 1.
 CN HMW1 OR MG312.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman K.V., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
 RA "The minimal gene complement of Mycoplasma genitalium."
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 721-847 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
 RA "A survey of the Mycoplasma genitalium genome by using random
 RT sequencing."
 RL J. Bacteriol. 175:7918-7930(1993).
 CC - FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
 CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC - SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
 CC MEMBRANE (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U39712; AAC71534.1;
 DR EMBL: U02261; AAD12527.1;
 DR TIGR: MG312;
 KW Cytadherence; Structural protein; Complete proteome.
 SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3D856 CRC64;
 Query Match 4.0%; Score 174; DB 1; Length 1139;
 Best Local Similarity 19.9%; Pred.No. 0.26;
 Matches 169; Conservative 117; Mismatches 305; Indels 260; Gaps 45;
 QY 70 YNGRVYDALFSELLMKDPNYOLK-DADIVNEVKGYYII---KVDGKYVYVYKDAHAD 125
 DB 27 YDGG-----STVEEDPNIKVAYDAD-----GNGYIAFNKGTGVYDYPYGDTEY-D 71
 QY 126 NVRTKDE-----INRQEHVKNEKNSVNAVARSQRYTTNDGYVFNADIIETGNA 180
 DB 72 ISQLFDENGPNFVDFEKOEE-NVLYKTVGN-----PDGYSDENGEWNWS----- 115
 QY 181 YIVPHGGHYHYIPKSDLSASLAAKHAHLAKNMQPSLSSTPSLPIPGTSHEKH 240
 DB 116 -----GVF-----END-----QWISTKESQPTDENYGF--DSDDL--PEVKOPES 151
 QY 241 EEDGYGFANRI-----IADESEF-----VMSHGDHNNHYFFKKDLTEE 279
 DB 152 VEDNYGFD-NDLPPPEVKOPESVEDNYGFDNDLPEVKOPESVVDQPSDDYFAKQP-TDE 209
 QY 280 QIKAAQKHLBVKVSHNGLOSLS-----SHEQDYPNSNAEMKDLQKIE 323
 DB 210 NYGFDNDLPPEVKOPESVVDQPSDDHFAKQPESTTDSYFSDSLDPTLDQPSLDHVV- 268
 QY 324 EKIAGIMKQGVKRESIVVKNKNAIYPCGDHHADPDEHPVGTGSHSNVLEFKPE 393
 DB 269 -----QYNF-----DHH-----EELKPV--AEQNNYQV----- 290
 QY 384 EGVAKKGNKYVTGEELTNVNNLLKNSTFNQNTFLANGQKRVYSFSPPELEKRLGINML 443
 DB 291 -GFDOVQAN-LDNEETOPTAEKVVITDFESKQAVVD-----SYQLPITDQDQQTFS 343

QY 444 VKLIT-PDGKLVKGVSGKVFGEVGNIANFELDQPLPGQTFKYTIASKDYPEYSDGTF 502
 DB 344 SSFQPTQVEQFDQVNSEV-----NQQKPEITKEPVLSESNFKQD 384
 QY 503 TVPTSLAYKMASQITFYFFHAGDTYLRVNPQFAPVKPQTDALRVF-----DEFH 551
 DB 385 VVETS---DLNSESNLXSENKDA-----TNDLSLNSFIQLNSNETASDDVH 430
 QY 552 GNAVLE--NNYKVGKILPIPKLNOGTTRTAGNKIPYTFMA-----NAYLDNQSTYI--VE 603
 DB 431 YESKSEPIHDYKFGS-----DLSQSNNSNLSSEPKFENSETAPDAHFESQSPVQVQ 484
 QY 604 VPILKEN---QTDKPSILPQFKRKAQENSKLDEK-----VEPKTSEKVEKE 649
 DB 485 YDIYQNEELKPTLDQPSDDYFAKOPTDENYGFNDLPPVEVKQPEVVDQPSDDHEAKQ 544
 QY 650 KLSETGNTSNTLEEYPTVDP-----VQ-----EKVAKFASISYGMKLENVLFNM 694
 DB 545 PESTTDSYFSFDLPO--PTLDQPSLDHVVQNFHDHHEELKPVAEQNNYQVGFQVQANL 603
 QY 695 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSKSTGTGVNQPTENKPADSLPE 754
 DB 604 DNNEIQ-PTAE--KEVTTDFESKQAQ-----VVDSYQLPIDT--DQDQDTFESSFET 652
 QY 755 APNEKPVKPESTNDGMLNPGNVGSDPMLDSALEAPVDPVQEKLEKTASYGLGLDS 814
 DB 653 QPTVQFQDVANSEVNDQFKPE--ITKEPVLESFNKQDVW-----TSNTNNLQK 701
 QY 815 VIFNMDGTIEL 825
 DB 702 FDIQSDNKITI 712
 RESULT 5
 CNA_STAAU STANDARD; PRT; 1183 AA.
 ID CNA_STAAU STANDARD; PRT; 1183 AA.
 AC Q53654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen adhesin precursor.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FDA 574;
 RX MEDLINE=92165839; PubMed=1311320;
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
 RA Lindberg M., Hoeoek M.;
 RT "Molecular characterization and expression of a gene encoding a
 RT Staphylococcus aureus collagen adhesin."
 RL J. Biol. Chem. 267:4766-4772(1992).
 RN [2]
 RP ERRATUM.
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
 RA Lindberg M., Hoeoek M.;
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC STRAIN=FDA 574;
 RX MEDLINE=94032261; PubMed=8218209;
 RA Patti J.M., Boles J.O., Hoeoek M.;
 RT "Identification and biochemical characterization of the ligand
 RT binding domain of the collagen adhesin from Staphylococcus aureus."
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
 RX MEDLINE=97475225; PubMed=9334749;
 RA Symersky J., Patti J.M., Carlson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,


```

RA  Narayana S.V.L.;
RT  "Structure of the collagen-binding domain from a Staphylococcus
RL  aureus adhesin.";
RL  Nat. Struct. Biol. 4:833-838(1997).
CC  -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC  COLLAGEN-CONTAINING SUBSTRATA.
CC  -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC  an amide bond (Potential).
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
DR  EMBL; M81736; AAA20874.1; -.
DR  PDB; 1AMX; 24-JUN-98.
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  TIGRfams; TIGR01167; LPXTG_anchor; 1.
DR  PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
KW  Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT  SIGNAL          1      29      POTENTIAL.
FT  CHAIN           30     1154    COLLAGEN ADHESIN.
FT  PROPEP         1155     1183    REMOVED BY SORTASE (POTENTIAL).
FT  DOMAIN         151      318    COLLAGEN-BINDING.
FT  DOMAIN         533     1093    3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT  DOMAIN         1093     1157    LYS/PRO-RICH (CELL WALL-SPANNING).
FT  REPEAT         533      719    B1.
FT  REPEAT         720      906    B2.
FT  REPEAT         907     1093    B3.
FT  SITE           1151     1155    LPXTG SORTING SIGNAL (POTENTIAL).
FT  MOD_RES        1154     1154    AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT  SEQUENCE       1183 AA; 133066 MW; 66ALCC07E575D76 CRC64;
Query Match
Best Local Similarity 3.9%; Score 173; DB 1; Length 1183;
Matches 197; Conservative 128; Mismatches 352; Indels 320; Gaps 48;
Qy  16  NRRYSYVDGSSQSKSENLT-PCDVSQKEGIAEQIVIKITDQGVYVTHGDDHYHYNGK- 73
Db  195  NNEKSYV-----SKDITIKDIOGGQQLDLSTLNIIVT-----GTHSYSGQS 238
Qy  74  --VPYDALFSLEMLKDP-----NYQLKDADIVAEVKGYYI--- 107
Db  239  AITDFEKAFPGSKITVONTKNTIDTIPQGVSYNSFSINKTK---ITNEQQKEFVNS 295
Qy  108  -----IKVDGKYVYLKDAHADNVR-----TKDEINRCKQEH----- 140
Db  296  QAWYQEHGKEVNGKSF---NHTVHNINAGIEGTVKGLKVLKQDKTKAPITANYKFK 352
Qy  141  -----VKDNEKXVNSNVAVERS-----QGRY----- 160
Db  353  LSKKDGSVKVDNQKEIEIITDANGIANIKALPSCGYILKIEAPRPTFPKDXEYPTMK 412
Qy  161  -TTNDGY--VFNPADIETD-----GNAYIVPHGGHYHYIPKSDLSASELAAGA 207
Db  413  DTDQGGFTTIENAKATEKTKDVSQAQWEGTQVKP--TIYFKLYKQDDNQNTTPVDKA 470
Qy  208  HLAGKNQPSOLSYSSPSPS-----LPINPGTSHKEHEDDGVGFANDRIIADESG 259
Db  471  EIKKLDGTTKYVWSNLPENDKNGKATKYLKVEKNAGGEDTTPGY----- 522
Qy  260  FVMGHGDNHYFFKKDLTEEQIKAAQKHLFEEV---KTSHNG-----LDSLSSHEH- 306
Db  523  LVVFN-----TERPIETTSISGEKVDKDNQDGKRPKXVSNLLANGEKVKVT 570
Qy  307  -DYPNNAK---EMKDLDKKTE-EKIGAIMQYGVK-----RESIVVNK-----EKNAIIV 351
Db  571  LDVTSETNWKYEFKDLPKYDEGKRIEYTVTDEHVKQDVTIDGTTITNKYTPGETSAITV 630
Qy  352  PHGDHHRADPIDEHKPYGIGHSHSNVLEFPPEGVA-----KK 389

```

369 QY --GIGH-- --SHS-- --NYELFPEGVAKKCGKNVYTGEBELTNVALLKN 401
119 Db QEASYGHDAPAHSVRTFTTSKDEKRDVPIHHPLSELSDRESRETHESLINTPVSLSG 178
410 QY STFNNQNT-----LANGKRVSVSFSP----- 431
179 Db TEDVTSTFAPSGDDGYLDGKKGVNVETPITLEESAVSDYLSGSVSNYQSKVDTPTKEETG 238
432 QY --PELEKGLGINMLVKLITPDGK-----VLEK-----V 457
239 Db GVPEIARSFG-NMEYTESPDQKGFQERDLSRSEKFEKFDQDFDSVLGKDSPAKFPGE 297
458 QY SKGVFGEVGVGNIANFELDQDYLPGQTFKTYIAASKDYPEVSDYDGTFTVPTSLAYKASOTI 517
298 Db SGWFFPVGFGDESAGLE-----KDPFTRSHDFDMKTETGMDTNSPSRSH 342
518 QY FYPFHAGDTYLRVNFQ-FAVPKGTDALRVFDFEFGNAYLENNYKVGEIKLPIPKNQCT 576
343 Db EFLDKTESGNDKNSPMGFGSESAG-LEKEFDQKNDG--RNEY'S-----PESDGL 391
577 QY TKTAGNKITPTFMANAYLDNQSYIYVEVI----- 606
392 Db GAPLGGNFPVR-SHELDLKNESIDKDVPTGDFGPDFLAKGRPGYGEASEDKFPARSD 450
607 QY -----LEKENQTD-----KPSLTPQFKKKAQENSKLDEKVEEPTSEKVEK- 648
451 Db DYEVEITELGRPKTETLDQFSPELSHPKSRDFEKESRDDFEETRDEKTEEPKQSTYTEKF 510
649 QY -EKLSETGNSTNSTLEVTVD-----PVQEKVAFASYSYGMKL-----E 688
511 Db ASMLGYSGEIPVGQDTQVAGTVDEKLTVPVNEKDOETESAVTTKLPISGGSGVVEEQRGED 570
689 QY NVLEFNMOGTIELYLPSEVITKKNMADFTGEAPQGENKPKPSNGKYSTGTVENOPTENKP 748
571 Db KVSQRDVIYAEKLITEEB--DKAFSDMVAEKLIGGEEBKE--TTTKEVEKISTE-KA 624
749 QY ADSLPEAPNEK-----PVKPEN-----STDNGMLNPEGNVGS 780
625 Db ASEEGEAEVEEVKGGGMVGRIGKWFGGGATDEVKPESPHSVVEEAPKSSGWFGGGATEEV 684
781 QY DMWLDALAEAP-----AVDPVQEKL 801
685 Db KPSKPSHSVEESPQSLGSTVVPVQKEL 710

RESULT 7
ID FNBA_STAAU STANDARD; PRT; 1018 AA.
AC DC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibronectin-binding protein precursor (FNBP).
GN FNBA.
OS Staphylococcus aureus.
CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 8325-4;
RX MEDLINE=89089996; PubMed=2521391;
RA Signaes C.; Raucci G., Joensson K., Lindren P. E.,
RT Anantharamiah G.M., Hoeck M., Lindberg M.;
RT "Nucleotide sequence of the gene for a fibronectin-binding protein
RT from Staphylococcus aureus: use of this peptide sequence in the
RT synthesis of biologically active peptides."
RL Proc. Natl. Acad. Sci. U.S.A. 86:699-703(1989).
CC !- FUNCTION: THE ABILITY OF BACTERIA TO BIND FIBRONECTIN HAS BEEN
CC PROPOSED AS A VIRULENCE FACTOR ENABLING BACTERIA TO COLONIZE
CC WOUND TISSUES AND BLOOD CLOTS. BINDING OF PLASMA FIBRONECTIN TO
CC THE BACTERIAL SURFACE MIGHT BLOCK ADHESION RECEPTORS ON S.AUREUS.
CC THIS REPRESENTING AN IMPORTANT DEFENSE MECHANISM AGAINST TISSUE
CC INVASION.

```
CC CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC CC an amide bond (potential).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; J04151; AAA26632.1; -.
CC DR InterPro; IPR004237; Fm_bind.
CC DR InterPro; IPR001899; Gram_pos_anchor.
CC DR Pfam; PF00746; Gram_pos_anchor; 1.
CC DR Pfam; PF02986; Fm_bind; 1.
CC DR TIGRfams; TIGR01167; LpxTG_anchor; 1.
CC DR TIGRfams; TIGR01168; YSIRK_signal; 1.
CC DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
CC KW Cell wall; Peptidoglycan-anchor; Repeat; Signal.
CC FT SIGNAL 1 36
CC FT CHAIN 37 985 FIBRONECTIN-BINDING PROTEIN.
CC FT PROPEP 986 1018 REMOVED BY SORTASE (POTENTIAL).
CC FT REPEAT 545 574 B-1.
CC FT REPEAT 575 604 B-2.
CC FT DOMAIN 745 878 4 X APPROXIMATE TANDEM REPEATS,
CC FT FIBRONECTIN-BINDING DOMAIN.
CC FT REPEAT 745 782 D-1.
CC FT REPEAT 783 820 D-2.
CC FT REPEAT 821 859 D-3.
CC FT REPEAT 860 878 D-4 (INCOMPLETE).
CC FT DOMAIN 879 948 5 X TANDEM REPEATS, PRO-RICH (WR).
CC FT REPEAT 879 892 WRI.
CC FT REPEAT 893 906 WR2.
CC FT REPEAT 907 920 WR3.
CC FT REPEAT 921 934 WR4.
CC FT REPEAT 935 948 WR5.
CC FT SITE 982 986 LPXTG SORTING SIGNAL (POTENTIAL).
CC FT MOD_RES 985 985 AMIDE-LINKED TO CELL WALL (POTENTIAL).
CC SQ SEQUENCE 1018 AA; 111780 MW; 58175E0020E81FLF CRC64;

Query Match 3.9%; Score 171; DB 1; Length 1018;
Best Local Similarity 19.4%; Pred.No 0.32;
Matches 169; Conservative 110; Mismatches 301; Indels 292; Gaps 43;

QY .11 QENKDNRVSYVDGSQSQSKSENLTPOVQSQKEGIOAEIVIKITDGYVTSHGDHYHY 70
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 313 QTNGNQITITSLABEQTSKELD-----VRYKDG|-----GNYYANL 348
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 71 NGKY----PYDALFSELLMKDPNYQLKDADIVNEV-----KGGYIIKVDGKYVYL-- 118
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 349 NSGIETFNKANRFHVAFIPNNGKTSTTVGTLMKSGNQNGNQBPVK--RIFEYLGNN 406
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 119 KDAAHADNVATKDEINRKQEHYKNDKVNKNVAVASQGRYTNDGVFPNPADIIEDTG 178
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 407 EDIAKSIVYANTD---TSKFKEVTSNMGSGLNL---QNNGSYSLN-----IENLD 450
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 179 NAYIVPGHGHHYIPKSDLSASELAANKAGLAGKNMOPSQL-----SYSSTPSPSLPI 231
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 451 KTYVVHVDGBY-----LNGTDEVDFRTQMGV---HPEQLYXYDYDRGYTLT----- 493
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 232 NPQTSHKEHEPDYGFDANRIADAESGFVMSHDHNHYFFKDLTBEOIKAA--OKHL-- 288
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 494 -----WDGNLVLYSNRANGNEKGPII---QNNKFYEKDTKETITLGGYDKNLVT 541
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 289 ---EEVKTS-----HNGLSLSLSEHDYPSNAKEMKD--LDKKIEPKIAGIKMQYGVKR 337
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 542 TVVEYDSSTLDIDYHTAIDGGGYVDGYETIETIEDSSAIDIDYHTAVDSEACHVGYYT 601
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 338 ESTVWNEKNNAIYPHGDH---HHADPIDEHKPVGIG---HSHSNYELFKPE--EGV- 386
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

Db 602 ES---SEESPIDFEESTHNSKHADVVEEDTNPGGCQVITESNLVFDEESTKGIIV 658
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
```

```
or send an email to license@isb-sib.ch).
CC EMBL: M87489, AAA24966.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGAL; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-K.
FT REPEAT 1109 1116 1.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2657807A6 CRC64;

Query Match
Best Local Similarity 18.8%; Score 168; DB 1; Length 1702;
Matches 175; Conservative 123; Mismatches 347; Indels 286; Gaps 42;
QY 12 ENKNNRVSYVDGSG--SSQKSNLTPDQVSGKEGIAQEQIKITDQGYVTSHGDIHY 70
DB 467 ENKSLKVG--DGTILKQADANNKAFQSGVIGRSTVWLNDKQVDPNSIYFGFR 524
QY 71 NGKVPYD-----ALFSEELMKDPN---YOLKDAI 98
DB 525 GRLDANGNLTFFHIRNIDDCARLVNHTSKTSVITIGESLITDPTIIPYNIDAPDE 584
QY 99 VN-----EVKGG---YIKVDGKYVYVLDAAHADNV-----RYKDEINR 135
DB 585 DNPYAFRIKDGQGLYLNLENTYVYALRGASTRSELPKNSGESNENLYMGKTSDEAKR 644
QY 136 QKQEHVKDKENSVNAVARSQRTTNDGVYFNPADIETDGNAYIVPHGGHYHIFKS 195
DB 645 NVNWHI--NERNWG-----FNGYGEEG-----KNGNUNVTFKG-----KS 681
QY 196 DLSASELAAKAHAKGNMPSQSYSTPSPSLPINTGSHKHEEDGYFDANRIAE 255
DB 682 EQNRFLLTGTTNGLDGLKVEKGTFLSGRPTPHARDIAGISSTK--KQHFENNEVVVE 739
QY 256 DESGFVMSGHDHNYFFKK--DLTEEQIKAAQKHLVEVKTSHGNDLSLSH----- 304
DB 740 D-----DWINRFRATNIVTNATLYSGRNVANITSNITASDNKAVHIGYKAGDT 790
QY 305 ---EODYPSNAKEMKDDKKIEKIAGIMKOYGVKRESIVNKEKNA----- 348
DB 791 VCVRSYDTGYVCTTD--KLSDK---ALNSFNATVSGNVNLGNANFLVLCANLFGTI 844
QY 349 -----IIPGCDHH--ADPIDEHKVPGLGHSNVELFPEEGVAKKGNKYVTGEE 399
DB 845 SGTGNSQVRLTENSHTLTCGDSNVQNLNDKGIHLN-----ACNDANKVTTTNT 894
QY 400 LTNVNLKKNSTFNNQNTLANGQKRVSFSPPELEKILGNMLVCLITPDGKVLKESVG 459
DB 895 LT--VNSLS-----GNG-----SFYLTDLNKGDKRVVTK--SATGNFTLQVAD 936
QY 460 KVFGEVGNIANFELDQPLPGQTFKYTTASKDYPEVSDGTFVTPTSLAYKMASQTIY 519
DB 937 KT--GEPTKN-----EL-----TLFDASNATRNUNLSLVGNTVDLGAWKKYLRNVNGRY 984
QY 520 PFHAGTYLRVNPQ-----FAPKGTDLVRVDFEFGHAGNVLNNKYKGEIKL 567
DB 985 -----DLY---NPEVEKRQNTVDITNITPNKIADV-----PSVPSNNEELARVET 1028
QY 568 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVVEVPILEKENOTDKPSILLPQKRNKA 627
DB 1029 PVPP-----PAPAT-----PSETTETV-----AENSKQESKTVKEKNEQDAT 1064
QY 528 QENSKLDEKVEEBPKTSEK--VEKEKLSETGNSTNSTLEEVFTVPVQ--EKVAKFAESYVG 684

Db 1065 ETTAQNGEAAEEAKPSKANTQNTNEVAQSGSEETETQTTEIRKTAKEKAKVEKEEK 1124
QY 685 MKLENVLNWDGTLIELYLPSEGEVIKKKNMADFTGEAPQNGENKRP-----SENGKV 734
Db 1125 AKVEK-----DEDEAFQMASETSPKQAKPAPKEVSTDTKV 1160
QY 735 STGTIVENQP-----TENKPADSL--PEAPNEKPVKP---ENSTNDGMLNPE 775
Db 1161 EETQVQAQPTQSTTVAABEATSPNKPAEETQPSKTNAPETVPVYSKNOTENTTDQPT 1220
QY 776 GNVGSDPMLDSALEEAPAV---DPVQEKLE 802
Db 1221 EREKTAKEVETKTOEPPQVQASQSPKQEOSE 1251

RESULT 9
MRSP_STAAU STANDARD; PRT; 1637 AA.
ID MRSP_STAAU STANDARD; PRT; 1637 AA.
AC P80544; Q9ZF62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Methicillin-resistant surface protein precursor.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 1061;
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "pls, a large repeat-rich surface protein of methicillin resistant Staphylococcus aureus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
RP 1199-1205 AND 1217-1224.
RC STRAIN=Isolate 1061;
EX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelae J., Vuoto M., Kuusela P.;
RT "Purification and characterization of a plasmin-sensitive surface protein of Staphylococcus aureus.";
RL Eur. J. Biochem. 236:904-910(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF115379; AAD09131.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
DR Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor; Repeat; Signal.
FT SIGNAL 1 48
FT CHAIN 49 1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT PROPEP 1602 1637 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-(SAG).
FT SITE 1598 1602 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1601 1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match
Best Local Similarity 20.2%; Score 165.5; DB 1; Length 1637;
Pred. No. 1.1;
```

```
Matches 168; Conservative 88; Mismatches 327; Indels 247; Gaps 39;
QY 120 DAHADNVTKEINRQKQEHVKD-----NEKVNNAVAVAR-----SOGRYTTNDGY 166
DB 323 EKAEIEKVLPKD-ISNLSNEELKKIALSEVLKETANKENAPRAFERSVSSNARTVNY 381
QY 167 VFN-----PADIIETGNAYIVPHGHHYIPKSDLSASEIAAAKAHLAKGNMQPS-----217
DB 382 SATALRAAAQDVTTKGTGNTFAHGDILHKTYKEEFPNEGTLTA-----FNTNFNPNITGK 437
QY 218 -----QLSYSTSPSPSLPINPGTSHKHEDEGYGFANDRIIARD-----ESGFVMSHGHD 267
DB 438 GALEYNDKIDFNKDTIIVPVANNQGTGADGNGFNTGONGDFLNQGGILRDKMA 497
QY 268 NYHFFKKLTESQIAAKOKHLEVTSTSHGLDLSHSEQDYPSNAKEMKDLDKKIEEXIA 327
DB 498 NASGEKIDTAYNNVAGVKDKLDADKT--NNLSQIGAAKVGYGVFKNGAD-----545
QY 328 GLMKGYGVKRESIVNKKK--NAILYPHGDHHDHAPIDIEKPVGIGHSHS-----NY 377
DB 546 GVTNVG---QNALVTKDKPVNKIIYADNTTNHLD-----GQFHGORLNDVNLNY 592
QY 378 ELFKPEEGYAKKGNKVTGEBLTNVNLLKNSTENNQNFTLANGOKRVSPS---FPPEL 434
DB 593 D-----AATSTITATYAGTKWAKATTDLGIDKSQKYNELITSSHMNRYNSNGIMRTNL 645
QY 435 EKKLGINMLVLLITPDGKVLKYSKVGSGVGNFIANTFELQOYLPQGTFFYTTASKDYP 494
DB 646 EG-----VLTTPQADLDDV--EVTKQPIPHKTIREFDPTLEPGS-----P 685
QY 495 EY--SYDG--TFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVFKG--TDALVRVF 547
DB 686 DVIVQKGEDGKTTTPT--KVDPDI-----GDVVERGEPTTEYTKNPVDEIVHFT 734
QY 548 -----DEPHGNAYLENNYKVEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQ 598
DB 735 PEEVPOGKHDEFDPN-----LPI-----DGTEEVPG-----KPGIKNPE 768
QY 599 TVIVEVPILK-----ENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKE 649
DB 769 TGEVVTTPVDVTKHGPAGPEVTKESIPPEKKREFNPDLPKEKV-----TQEGQTGE 824
QY 650 KLSGTGNTSTLSEV-----PTVDVQVEKVAFAESVGMKLE-----NVLFNMDGTI 698
DB 825 KTTTPTTINPLTGKVGEGEPTTEVTKEPVDEITQFGEEVPOGKHDEFDPNLPIDGTE 884
QY 699 ELYLP-----SSEVIKKNMADFTGEAPQGN-----GEN 726
DB 885 E--VPEKFGIKNPETGEVVTTPVDVTKHGPAGEPEVTKESIPPEKKREFNPDLPKGE 942
QY 727 KPSNGKYSTGTV-----ENQPT-----ENKPADSLPEAPNEX-----759
DB 943 KVTQSGQTGKTTTPTTINPLTGKVGEGEPTTEVTKEPVDEITQFGEEVPOGKHDEF 1002
QY 760 PVKPENSTDN-----GMLNPE-GNVGSDPMDLSALEAPAVDPVQEKLE 802
DB 1003 DPNLPIDGTEVPGKPGIKNPETGEVVTTPVDVTKHGPAGEPEVTKEE 1052

RESULT 10
IGAL_HAEIN
ID IGAL_HAEIN STANDARD; PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-JUN-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
```

```
RP
RC STRAIN=HK368 / Serotype B:
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.:
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RL of Haemophilus influenzae serotype b.";
RN Infect. Immun. 57:3097-3105(1989).
[2]
RP
RC STRAIN=HK368 / Serotype B:
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.:
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases.";
RN J. Bacteriol. 174:2913-2921(1992).
CC
CC -I- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC
CC -I- SUBCELLULAR LOCATION: Secreted.
CC
CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X64357; AAA45708.1;
DR EMBL: M87492; AAA24969.1;
DR MEROPS: S06.001;
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT MUTAGEN 288 288 S->T: LOSS OF ACTIVITY.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;
```

Query Match 3.88; Score 165; DB 1; Length 1541;

Best Local Similarity 19.3%; Pred. No. 1.1;

Matches 172; Conservative 120; Mismatches 331; Indels 268; Gaps 44;

```
QY 12 ENKDNRRVSVVDGSSQSKSGENTPDQVSKQEGIOAIOIVIKITDQGYVTSHGHHYIN 71
DB 462 DNKGLKVG--DGTVILKQQTNGSGQAFASVGLVSGRSTLVLNDDKQVDPNIIYGFRC 519
QY 72 GKVPYDA-----LFSEELMKDPN----YOLKQDADIV 99
DB 520 GRLLDNGNSLTFDHIRNIDDGARLVNHNMTNASNITITGESLITDPTTIPYNIDAPDE 579
QY 100 N-----EVKGG--YIIKVDGKYVYLKDAHADNV-----RTPKDEINRQ 136
DB 580 NPATAFRIKGGQGLYLNLENTYIALKKGASTRSELPKKNKSGESNENWLYMKYTSDEAKRN 639
QY 137 KQEHVKDNKVN--NVAVARSQGRYTTNDGYVFNPADIETDGNAYIVPHGHHYIPKS 195
DB 640 VMNHI--NNERMNGFNGVGECKNGNGLNVTFKG---KSEQNRLLTAGGTNLN---G 690
QY 196 DLSASELAAKAHLAKGNMQPSQLSYSTSPSLPINPGTSHKHEDEGYGFANDRIIAE 255
```

```
CC CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to licenses@isb-sib.ch).  
CC -----  
DR EMBL; M87490; AAA24967.1; -;  
DR MEROPS; S06.001; -;  
DR InterPro; IPR000710; IgA_S6.  
DR InterPro; IPR004899; Pertact_sup.  
DR Pfam; PF02395; IGAI; 1.  
DR Pfam; PF03212; Pertactin; 2.  
DR PRINTS; PR00921; IGASRPRTASE.  
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.  
FT SIGNAL 1 25 POTENTIAL  
FT CHAIN 26 1012 IMMUNOGLOBULIN AL PROTEASE.  
FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).  
FT ACT_SITE 292 292 PROBABLE.  
FT SEQUENCE 1545 AA; 170627 MW; 3EDD75398F6D478 CRC64;  
-----  
Query Match 3.7%; Score 164; DB 1; Length 1545;  
Best Local Similarity 18.6%; Pred No. 1.2;  
Matches 169; Conservative 123; Mismatches 312; Indels 306; Gaps 43;  
  
QY 12 ENKONRRVSYVDGSSOKSENLPDVQVSQKEIQAEIVIKITDGQYTSHGDDHYHYN 71  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 466 DNKGSLKYGV--DGTVILKQQTNGSQGHAFASVGISGRSTVLNDDKQDPNSLVFGFRG 523  
QY 72 GKVPYD-----ALFSEELLMKDPN---YQLKDAD-- 97  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 524 GRDLUNGSLTFDHTRNTDEGARLVNHNISTSKHSVTITGNNLTDPNNYSIYVRPLEDD 583  
QY 98 ---IYNEVKGGVII---KVDGKYVYVYLKDAHADNVET-----KDE 132  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 584 NPVATRIQIKYGQLYFNENETYYIALAKDA---SIRSEFPQNRGESNNWSLYMGTEKAD 639  
QY 133 INRQKEHVKNDEKYNVNVARSGRYTTNDGYVFNPADLIEDTGNAVTPVHGCHYHI 192  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 640 AQKNAMNHFI--NNERNWG-----FNGIFGEPEG-----KNGNLNVTFPG----- 677  
QY 193 PKSDLASALEAAKAHLAGKNMQESQSLSYSTPSPLPNPGTSHKEHEEDGYGFDPANRI 252  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 678 -KSEQNRELLTGCTNLNGDLAVQQOCTFLSGRTPPHARDIAGISSTK--KDSHFSENNEV 734  
QY 253 IADESGFVMSHGDNHHFFKKDLTEQIIKAQKHLEEVKTSHG--LDSSLSSH----- 304  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 735 VWBD-----DWTRNFKATNIVTNNTATLSGRNVESITSNITASN 776  
QY 305 -----EQDY-----PSNAEMKOLDKKI-----EE 324  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 777 AKVHIGYKAGDIVCRSDYGVTVCTDKLSDKALNSFPNLRGNVLTESANFVLGRA 836  
QY 325 KIAGIKWQYGVKRRESLVNVEKENALIYPHGDDHADIPIDEHKPVGIHGSNVELFKPEE 384  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 837 NLFCTIQSRGNSQVRLTENSWHW--LTGNSOVHOLD-----LANGHILN----- 879  
QY 385 GVAKKEGNKVYTGEELTNVNVLNKLNSTENNQNFTLANGQRKVSFSFPPELKKLGINMLV 444  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 880 -----SADSNNNVTKYNTL-TVNSLSGNG-----SFYILTDSLNRQGDKVVV 920  
QY 445 KLITPDCKVLEKRVSGKFEGGVGNITANELQPYPGQTFKYTTASKDYFEVSYDGTFTV 504  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 921 TK-SATGNTFLTQVADKT-GEPNHN----EL-----TLFDASKAQRDHLNVLGVNTVD 967  
QY 505 PTSIAVKMASQTIYFPFHAGDTYLRVNPQF-----APPKGTD 541  
DB :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
```

DB 968 LGAWKYKLVNGRY-----DLY---NPEVEKRNQTVDTTITTTNNIQAQVSPSPSNE 1019
QY 542 ALVRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNDQSTYI 601
DB 1020 ETARV-DE-----APVPPAPATPSETTETVAEN-----SKQESKT 1054
QY 602 VEVPILEKENOTD-----KPSILPQFKRK-AQENSKLDE---KVPEPK 641
DB 1055 VE-----KNQDATTTAQNRVAKKSNVKANTQTNVAQSGSETKETQTETRETA 1108
QY 642 TSEKVEKELSTGNTSGNSTLEEV-----TVDPVQEKVAKFAESYGMKLVNLFMDGT 697
DB 1109 TVEKEEAKV-ETEXT-----QEVPKVTSQVSPKQEQ-----SETVQQAEPAREN-DPT 1156
QY 698 IELYLPSGEVTKKNMADFTGAPOGNGENKPSNGKYSTG--TVENOPTENKPADSLPEA 755
DB 1157 VNIKEPQSQNTIADTEQAPAKETSSNVEQPVTESTVNTGNSVVEN-PENTTPATTOPTV 1215
QY 756 PNEKPKVDEN 765
DB 1216 NSESSNKKPKN 1225
RESULT 12
ID IGAO_HAEIN STANDARD PRT: 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0950.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59800; NOT_ANNOTATED_CDS.
DR EMBL: U32779; AAC22651.1; .
DR MEROPS: S06.001; .
DR TIGR: HI0990; .
DR InterPro: IPR00710; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;
Query Match 3.7%; Score 163.5; DB 1; Length 1694;
Best Local Similarity 19.0%; Pred. No. 1.5;
Matches 180; Conservative 125; Mismatches 385; Indels 259; Gaps 42;
QY 12 ENKDNREVSVDGSG-SSQKSENLTDPQVSKQEGIAEQIVIKITDQGYVTSHGDRHYHY 70
DB 467 ENKSLKVG--DGTVLKQQAQDANKVKAFSGVSGRSTVVLNDDQVDPNSIYGFGR 524
QY 71 NGKVPYD-----ALFSEELMKDPN----YQLKDADI 98
DB 525 GGRLDANGNNLTFEHRIIDGRLVNHNTSKTSTVITIGESLITDPTTIPNIDAPDE 584
QY 99 VN-----EVKGG---YLIKVDGKYVYVKDAHADNV-----RTKDEINR 135
DB 585 DNPYAFERRIKDGGQLYLNLENTYIALRKAGSTRSELPKNSGESNENLWYMGTSDEAKR 644
QY 136 QKQEHVXDNEKVSNAVARSQGRYTINDGVFNPAIDEDTGNAYIVPHGGHYHPKS 195
DB 645 NVNHHI--NERNMG-----FNGYFGEERG-----KNNGLNVTFKG-----KS 681
QY 196 DLSASELAATAKAGLAGNMQPSQISVSTSPSLPINPSTSHKHEEDGYGFDANRIAE 255
DB 682 EQNREFLTGTNLNGDLKVEKGTFLSGRTPHARDIAGISSTK--KDQHFAENNEVVVE 739
QY 256 DESGFVMSHGDNHYFFKK--DLTEQIKAAQKHLEEVKTSHNGLDLSLSSH----- 304
DB 740 D-----DWINRNFKATNINVTNNATLYSGRNVANITSNTASDNKAVHIGYKAGDT 790
QY 305 ---EQDYPSNAKEMKDLCKKIEEKIAGIKQYGVKRESIVVNEKNA----- 348
DB 791 VCVRSDDYGVVCTTD--KLSDK---ALNSEFATNVSGNVNLSGNANFVGLKRALFGTI 844
QY 349 -----IIVPHGDHHH--ADPIDEHPVQIGHSHSNVLELFKEEGVAKKGVITGEE 399
DB 845 SGTNSQVRUTENSHWLTGDSNVNQNLNDKGHIHLN-----AQDANKKVTYNT 894
QY 400 LTVNVLNKNSTFNQNTLANQKRVSFSPPELEKLGINMLVKLITPDGKVLKESVG 459
DB 895 LT--VNSLS-----GNG-----SFYLDLSNKKQDKVVVK-SATGNFTLOVAD 936
QY 460 KVFEGGVGNIANELDQPYLPGGTFKYTIASKDYPEVSGDTFTVPTSLAYKNASOTIFY 519
DB 937 KT-GEPTKN----EL-----TLFDASNATRNLNLSVGNITVDLGAWKYKLRNVNGRY 984

```
QY 520 PFAHAGTYLRVNPQ-----FAVPKGTDALRVFDEFHGNAYLENNYKVGEIKL 567
Db 985 -----DLY--NPEVEKRNQVDTNITTPNIOADV-----PSVPSNNEELIARVET 1028
QY 568 PIPKLNQGTTRAGNKIPVTFMANAYLNDQSTYIIVEVPILEKENOTD-----RPSILPQ 621
Db 1029 PVPVPAPAPP-----SETTETVAENSKOESKTVKEKNEQDATETTATAGQGEVABEAKPSVKAN 1084
QY 622 FKRNK-AQENSKLDE-KVEEPTSEKVEKELSETGNSSTLSEVPTV---DPVCEK 675
Db 1085 TOTNEVAQSGSTTEOTTEKETAKVEKEKAK-----VEKDEIQEAPOMASETSPKOAK 1140
QY 676 VAKFAESYGMKLENNLFNMDGTIELYPSGEVIKKNMADFTGEAPQNGENKPSN-----731
Db 1141 PAPKEVSTDKVEETQVQAQ-----POTQSTTVAAAEATSPNSKPAEETOPSEKTNAE 1193
QY 732 -----GKYSTGTVENOPTENKPADLPEAPNEKPKPENSTNDGMLNPEGNVGSDDPMLD 785
Db 1194 PVPVVKNOTENTTDQPTER---EKIAKVETEKTOEPPQVVASQSPKQOSETVQPOAV 1250
QY 786 SALEEPAPV---DPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGE 831
Db 1251 LSENNYPTVNAEEVQAQLOQTQTS-----TVSTKQPAPE 1285

RESULT 13
STC2_STA00 STANDARD: PRT: 715 AA.
AC P17855;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Staphylococcal precursor.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=BB.
RX MEDLINE=88139269. PubMed=3481366;
RA Kaiba S., Miyata T., Yoshizawa Y., Kawabata S., Morita T.,
RA Igarashi H., Iwanaga S.;
RT "Nucleotide sequence of the staphylococcalase gene: its unique COOH-
RT terminal 8 tandem repeats."
RL J. Biochem. 102:1177-1186(1987).
CC -!- FUNCTION: STAPHYLOCOAGULASE IS AN EXTRACELLULAR PROTEIN WHICH
CC SPECIFICALLY FORMS A COMPLEX WITH HUMAN PROTHROMBIN. THIS COMPLEX
CC NAMED STAPHYLOTHROMBIN CAN CLOT FIBRINOGEN WITHOUT ANY PROTEOLYTIC
CC CLEAVAGE OF PROTHROMBIN.
CC -!- DOMAIN: THE C-TERMINAL TANDEM REPEATS ARE NOT REQUIRED FOR THE
CC PROCOAGULANT ACTIVITY.
CC -!- SIMILARITY: TO THE STAPHYLOCOAGULASE FROM STRAIN 213.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
Best Local Similarity 20.5%; Pred. No. 0.51;
Matches 154; Conservative 82; Mismatches 253; Indels 263; Gaps 33;
QY 124 ADNVRTKDEINROKQEHVKDNEKVNNSVAVARQGRVTTNDGYVFNPAIDIEDTGNAYIV 183
Db 24 ADAIVTKD---YSKESRVNNSYGLTISDWYDKRLTSLSEOFINALDILET-----73
QY 184 PHGCHYHYIIPKSDLSASELAAKAHLAGKNMQPSQSYSTSPSPSLPINPGTSHKHEED 243
Db 74 -----YHY-----GEKEYKD 83
QY 244 GYGFADNRIIAEDSGVMSHGDHNYFFKDLITERQIKAAQKHLEEVTS-----294
Db 84 AKDKLMTIRILGEDQ--YLLER-----KKQYEEYKLYQYKEENPTSGKLKLTED 133
QY 295 -----HNGI--DSLSSHEDQYPSNAK-----EMKDLDKKIEEKIAGIM-----330
Db 134 QYTIEDITMREYNELTESLAKSAVKDEKVEKIENQHHDLDKPTDEMEKATSRVDDLAN 193
QY 331 KQYQV-----KRESIVNKEKNAILIYPHGDHHDADPIDEHKPVGVIGHSHSNYELF 380
Db 194 KAVSYVFAFVRDTHKTEALELAKAVDLVLG-----DEKPHRISNERIEKMI 242
QY 381 KPEGVAKKESGNKVTYTGELTNNVLLKNSTENNQFTLANGOKRVYSFPPPELEKKLGI 440
Db 243 KLESII--EDFIETGLNKPGNITSYSSKHHYKNS-----EGF 281
QY 441 NMLVKLITPDGKLVKSGVKGEGVGNIANFELDOPYLPQGTFTKYTIASKDYEVSYDG 500
Db 282 EALVKETR-----EAVAN-----ADESWKTKTVKYGSETSKSPVVKEN 321
QY 501 TFTVPTSLAYKMASOTIFYFFHAGDTYLRV-NQFAVPKG--TDALVRVDFEHGNAY--555
Db 322 KVEDPQPKEDNQOEKVTAGKAEETQVQAQPLVKIQGTITGEIVK-----GPEIPT 375
QY 556 LENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYL-DNQSTYIIVEVPILE-KENOT 613
Db 376 MENKTQGEI-----VQGDPTMEQSGPSLSDNYTQPTTPNPILLEGGSS 422
QY 614 DKPSILPQFRNKQAQ-----ENSKLDEKVEEPTSE-----KVEKEKLESG 655
Db 423 SKUEIKPQGTSTLKGQESSQIEVKPQATEATEASQYGRPOFNKTKPKYKVRDAGTG 482
QY 656 NSTSNS---TLEEVPTVDPVQEVKAFKFAESYGMKLENVLFNMMDGTIELXLPSEVKKNM 712
Db 483 IREYNDGTGFEYAPRPNKPSSETNA-----YNNVTNODGTVSY---GARPTQNK 528
QY 713 ADFTGE---APQNG-----ENKPS-----NGKVSSTGTVENQPTENKPAD 750
Db 529 ASETNAYNVTHANGQVSYGARPTQKPKSETNAYNVTHANGQVSYGA---RPTYNKPS 585
QY 751 SLPEAPNEKPKPENSTNDGMLNPEGNVGSDDP 782
Db 586 T-----NAYNVITHG--NQVSYGARP 605

RESULT 14
SC16_YEAST STANDARD: PRT: 2195 AA.
AC P48415; Q02822;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Multidomain vesicle coat protein.
GN SEC16 OR YPL085W OR LPF1W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017704; PubMed=7593161;
RA Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
```


RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that
interacts with Sec23p".
J. Cell Biol. 131:311-324 (1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Araujo R., Aparicio A., Barrall B.G., Badcock K., Allen E., Ansorge W.,
Bussey H., Storms R.K., Ahmed A., Albertman K., Benes V.,
Bostein D., Bowman S., Bruckner J., Carpenter J., Cherry J.M.,
Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
Hunnicke-Smith S., Hyman R., Johnston M., Kallman S., Kleine K.,
Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharif M.,
Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
Zhong W.W., Zollner A., Vo D.H., Hani J.,
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RT Nature 387:103-105(1997).
RL NATURE INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE
CC ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
CC CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION
CC CAUSES A LETHAL SECRETION DEFECT.
CC -1- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
CC WHICH BUD FROM IT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23819; AAC49088.1; -
CC EMBL: U41849; AAB68254.1; -
CC SGD: S0006006; SEC16
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094
FT MUTAGEN 1059 1059 L->S: IN SEC16-4; TS ACCUMULATION OF ER
FT MUTAGEN 1084 1084 L->P: IN SEC16-3; TS ACCUMULATION OF ER
FT MUTAGEN 1089 1089 L->P: IN SEC16-2; TS ACCUMULATION OF ER
FT MUTAGEN 1231 1231 W->R: IN SEC16-1; TS ACCUMULATION OF ER
FT CONFLICT 522 522 MISSING (IN REF. 1).
FT CONFLICT 560 560 I -> F (IN REF. 1).
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEBF0 CRC64;
Query Match 3.6%; Score 159.5; DB 1; Length 2195;
Best local Similarity 21.5%; Pred. No. 3.3;
Matches 193; Conservative 97; Mismatches 347; Indels 261; Gaps 48;
QY 13 NKDNRVSVSDGSQSSQSENITPDQVSOKEIQAEQIVKITDQGYVTS-----HGDRYH 68
DB 1426 NKKNDKKVDFDFTGSSANSSTVDLTOTFTFPFQAQ-----VTSQSYVDTTALLHNAH-- 1478
QY 69 YNKGVPYDALSEBELMKDPNQLKADIVNEVKGVIYKVDGYVYLLKDA----- 121
DB 1479 ----NVP-----SHSVLHKSPPS-----NVSKG--LVEANLPYTHRIGDSLQGSQPR 1518
QY 122 -----AHADNVRTKDEINRCK-----QEHVKDNEKYNVS 149
DB 1519 IHNTOFAAEAPOMASLRRVTRTDQHTNEKALSKSQILEKKSQTAYTPTQFGNHSVPMEKNS 1578

QY 150 NV-----AVARSQGYTTNDGCVENP-----ADITDTGNVAVIVPHGGHYHVPKSDLS 198
DB 1579 NVPSLFAFPAPKPLGTVPNS--YVSPDLVRRESITSGSELP-------PAIGVP 1628
QY 199 ASELAATAKAHLAGNMQPSQSYSSSTPS--PSLPINFTS--HEKHEEDGYGFDANRIAE 255
DB 1629 TK-----ANSQSGSLMYS--PSVEALPIDVPVPOVHETGYNDFGNKHKSQSMPE 1675
QY 256 DES--GFVMSHGSHNHFFPKDLTETEQIKRAAQKHLVEVKTSHNGLSLSHEDQYPSNAK 313
DB 1676 DESHTSHDNNAQONTLKSDADYTDMDTIEGFGFNDVK---NLLPMPNPHQSTSTVNPI 1732
QY 314 EMKDLOKIEEKIAGIMQGVARESTIVNKEKNALIIYPHGDHHDADPIDEHFKPVGIGHS 373
DB 1733 Q-----TISDDTQPL-QINVEVRGTDASKMENS--PSIENERS--SEEQENISKS 1780
QY 374 HSNVLPKPEGVAKREGNVYVYGEELTNVNVLLKKNSTENNQNFTLANGOKRVSFPEPPE 433
DB 1781 ASS--AYLPSTGSLSE-NRPLTQDE--NSISSETGVSTY-----LPAG----- 1818
QY 434 LEKLGINMLVKLITPDGKVLKVGKVGFGVGNIA-----NFELDDQPYLPG 481
DB 1819 -----SISMEAKPLSQVDVPRNVNKKASKLVQGHAPPKPKSTDATKNY---SPYVP- 1869
QY 482 QTFKTYTIASKDYPEVSYDGTFTVPTSLAYKMASQ-----TIYPF--HAGDTYLVNPO 533
DB 1870 ---QSTAAASADGDE---SHILKTSPIAYIARTHQAHASNPSEYFPLVNOANET-----AS 1917
QY 534 FAVPKGTDALVRVDFEFGHAYLENNYKVGGEIKLPIKLN-----QGTTRTAGNK--- 583
DB 1918 FESESTSA-----QSGNGVASENRF-----PIKAEVVEKDTFOPTIRKASTNOYR 1966
QY 584 --IPVTMANAYL-----DNQSTYIIVEFILEKENQTDKPSILPOFKRKAQENSK 632
DB 1967 AFKPLESDADKNDVIEDSDDDNMSTDEAKNRKEKKNVMMKTKP-----SNKD 2018
QY 633 LDEK-----VEEPKTSKVEKEKLETSNGTSNSTLEEVPTVDPQVKVAFESYGM 685
DB 2019 IDDSNGWFWLKKDGTGKKYKAKLGHKNTLYYDEKLRKWNKDKATEEBEKQKLISSAP 2078
QY 686 KLENVLFNMD-----GTIELYLP-----SGEVIKKNMADFTGE-----APQG-NG 724
DB 2079 PPPIVRKKDGGPKTKPRSGPINSLPPVHATSVIPNN--PITGEPLPIKTSPTSPTGPNP 2136
QY 725 ENKPSSENGKVS--TGTVENQFTENKPADSIPEAPNEKP--VKPENSTDNGLNPEGNV 778
DB 2137 NNSPSPSPISRTSGVNLTSKKANGLDLILSLAGGPKPASTRRKKTKTARGYVNVMDNI 2194
RESULT 15
MSPL_PIAFM STANDARD; PRT; 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT Plasmodium falciparum";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 37.3021 Seconds

(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAYALNQHRSQENKONNRVS.....GTIELRLPSGEVKKNLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
- 1: sp-archaea.*
 - 2: sp-bacteria.*
 - 3: sp-fungi.*
 - 4: sp-human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp-organelle.*
 - 9: sp-phage.*
 - 10: sp-plant.*
 - 11: sp-rodent.*
 - 12: sp-virus.*
 - 13: sp-vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp-virus.*
 - 16: sp-bacteriap.*
 - 17: sp-archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4274.5	97.2	1039	Q9ANY1	Q9any1 streptococc
2	1002.5	22.8	844	Q9AG74	Q9ag74 streptococc
3	933	22.6	819	Q9ANY3	Q9any3 streptococc
4	989	22.5	819	Q97QM9	Q97qm9 streptococc
5	987	22.5	839	Q9ANY2	Q9any2 streptococc
6	975.5	22.2	802	Q97QM8	Q97qm8 streptococc
7	975.5	22.2	816	Q9AHT9	Q9aht9 streptococc
8	867	19.7	825	Q93GT5	Q93gt5 streptococc
9	864	19.7	825	Q99XV4	Q99xv4 streptococc
10	855.5	19.5	822	Q92HG7	Q92hg7 streptococc
11	689	15.7	289	Q9AE21	Q9ae21 streptococc
12	224.5	5.1	792	Q99Z76	Q99z76 streptococc
13	212.5	4.8	961	Q99RD3	Q99rd3 streptococc
14	210.5	4.8	2004	Q97QP7	Q97qp7 streptococc
15	208	4.7	1078	Q963T1	Q963tl plasmodium
16	196.5	4.5	940	Q53682	Q53682 staphylococ

17	194.5	4.4	1964	2	Q59947	Q59947 streptococc
18	194.5	4.4	2485	5	Q96134	Q96134 plasmodium
19	194	4.4	1139	5	Q97237	Q97237 plasmodium
20	191.5	4.4	1939	5	Q25662	Q25662 plasmodium
21	189	4.3	2269	5	Q26223	Q26223 plasmodium
22	189	4.3	2276	2	Q93TV6	Q93ty6 staphylococ
23	188.5	4.3	5458	5	Q9U459	Q9u459 plasmodium
24	187	4.3	1873	2	Q9ZAN7	Q9zan7 enterococcu
25	186	4.2	1524	10	Q8RYN2	Q8ryn2 oryza sativ
26	183	4.2	1236	5	Q9GTX2	Q9gtx2 plasmodium
27	182	4.1	1475	5	Q25842	Q25842 plasmodium
28	180	4.1	1231	2	P72362	P72362 streptococc
29	179	4.1	1650	5	Q77328	Q77328 plasmodium
30	177.5	4.0	1271	5	Q25860	Q25860 plasmodium
31	177	4.0	2747	5	Q9BJX9	Q9bjx9 plasmodium
32	177	4.0	2910	10	Q9RND5	Q9rns5 arabidopsis
33	176	4.0	2481	16	Q99RG6	Q99rq6 staphylococ
34	175.5	4.0	940	16	Q928N7	Q928n7 listeria in
35	175.5	4.0	1134	2	Q99051	Q99051 streptococc
36	174.5	4.0	1435	5	Q9NG63	Q9ng63 plasmodium
37	173.5	3.9	786	5	Q77357	Q77357 plasmodium
38	173.5	3.9	1141	6	O46489	O46489 galago cras
39	173.5	3.9	1576	13	Q9I908	Q9i908 xenopus lae
40	173.5	3.9	1927	2	Q54875	Q54875 streptococc
41	173	3.9	2478	2	Q9RL69	Q9rl69 staphylococ
42	172.5	3.9	3484	5	P91257	P91257 caenorhabdi
43	172	3.9	710	10	Q39060	Q39060 arabidopsis
44	172	3.9	2478	2	Q9LCH2	Q9lch2 staphylococ
45	172	3.9	4688	16	Q9PQ08	Q9pq08 ureaplasma

ALIGNMENTS

RESULT 1

Q9ANY1	PRELIMINARY;	PRT; 1039 AA.
ID	Q9ANY1	
AC	Q9ANY1;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)	
DE	Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).	
DE	SP1004.	
GN	PHIE OR SP1004.	
OS	Streptococcus pneumoniae.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;	
OC	Streptococcaceae; Streptococcus.	
OX	NCBI_TaxID=1313;	
RX	SEQUENCE FROM N.A.	
RP	MEDLINE=21101045; PubMed=11159990;	
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,	
RA	Dormitzer M., Dagan K., Brewah Y.A., Barren P., Lathigra R.,	
RA	Langermann S., Koenig S., Johnson S.,	
RT	"Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis."	
RT	Infect. Immun. 69:949-958(2001).	
RL	[2].	
RL	SEQUENCE FROM N.A.	
RP	STRAIN-TIGR4;	
RC	MEDLINE=21357209; PubMed=11463916;	
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,	
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,	
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,	
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,	
RA	Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,	
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,	
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,	
RT	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."	
RT	Science 293:498-506(2001).	
RL	Science 293:498-506(2001).	
DR	EMBL; AF318956; AAK06761.1;	


```

Db 441 KOESLHKLGAKKTDLPSSDREFY-----NKAYDLARIHQDLDLNDKGRQVDEEA- 490
QY 564 EIKLPKPKLNOGTRPAGNKIPYTMANAYLDNQSYIVEVPILEKENQTDKPSILPOFK 623
Db 491 -----LDNLELKVSSDKVLVDIIAFL-----APIRPE-RLGKPNQAIIYT 535
QY 624 RNKAQNSKL-----DEKVEEPK-TSEK-----VEKEKSETGNSTSN 660
Db 536 DDEIQV-AKLAKGYTTEDGYIFDPDITSDEGDVYTPHMTSHHWIKKDSLSEARAAQ 594
QY 661 SILEEVPYDPVQE-----KYAKAEYS-----GKMLNVLFNMDGTEIYLPGE 706
Db 595 AVAKERGLTPPTDQDSNGTEAKGAELNRYKAAKVPLDRMPYNLQYTV-----VKNGS 652
QY 707 VIKKNMADFTG-----EAPQ-----NGENKPSE- 730
Db 653 LIIPHYDHYHNKKEWFDGLYAPKGYSLDILLATVXYVYHPNRPDSNGFNGASDH 712
QY 731 -----NGKYSTVTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVGSDPMLD 785
Db 713 VORNGAQADTQTE-KPNEEKPOTEKPEETPREKQSE-----KPSPKPTPEEP 765
QY 786 SALEEAP-----AVDPQVQKLEK 803
Db 766 ESPSPSESEBPQVETEKVKLEKRE 791

RESULT 3
Q97QM9 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.
AC Q9ANY3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Pneumococcal histidine triad protein B precursor (Fragment).
GN PHTB.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINP=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
DR EMBL; AF318954; AAK06759.1;
KW Signal.
FT SIGNAL.
FT NON_TER.
SQ SEQUENCE 819 AA; 92108 MW; E502FC16CC28A5F CRC64;

```

```

Query Match 22.6%; Score 993; DB 2; Length 819;
Best Local Similarity 31.8%; Pred. No. 9.4e-40;
Matches 297; Conservative 120; Mismatches 263; Indels 254; Gaps 36;

QY 1 CAYALNQHRS-QENKDNRYVYDGSQSQSENLTDPVQSKGIGIAEQIVIKITDQGY 59
Db 20 CSYELGRYOAGODKESNRVAYIDGDQAGAKAENLTDPVSKREGINAQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGKVPYDALFSEELMLKDPNVLKADIVNEKGGYLIKVDGYVYVYVK 119
Db 80 VTSBGDHYHYNGKVPYDAIISEELMLKDPNVLKADIVNEKGGYVYVYVYVK 139
QY 120 DAAHADNVRTDEINRQKQEHVKD-NEKVNNSVAVARSQGRYTTNDGVFNPAIIEDTG 178
Db 140 DAAHADNIRTKEEIKRQQRSHNSRADNAVAAARAGQRYTTDDGYIFNASDIIEDTG 199
QY 179 NAYIVPHGGHYHYIPKSDLSASELAAKAHLAKG--NMQPSQLSYSTPSP-----SL 229

```

```

Db 200 DAYIVPHGDHWHYIPKNSLSASELAAAEAYWNGKQGRPSRSSSSSYNANPAQPRLESENHL 259
QY 230 PINGTSH-----EKH-EEDGYGFDANRIIAEDESFGVMHGDNH 268
Db 260 TVTP-TYHQNOGENISSILLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGNHY 318
QY 269 HYFFKKDLTEBQIKAAQKHLEEV-----KTSNGLDSL-----SSHEDQVPSN 311
Db 319 HF-----IPYQSELEKRIARIIPLRYSNHWVPDSRPEEPSPOPTPEPSPQAPSN 373
QY 312 ---AKEMKDDKKIEEKIAGIMKQYGVKRESIVVNKEKNAIYIPGHDDHADPIDEKPV 368
Db 374 PIDGKLKVEARVKVD--GYVFEENGYSR-----YIPAKDLSAETAA 413
QY 369 GIGSHSNYELFKPEEGVAKKE-----GNKY-----TGEELTN 402
Db 414 GIDSKLAKQESLSHKLTKTKTDLPSDRFYNKAYDALLARIHQDLDLNDKGRQVDFEALDN 473
QY 403 VVNLKKNSTFNQNFILANGQKRVSFSPPELEKKLGINMLVKLITPDGKVLKRVSKVF 462
Db 474 LLERLKOVSSDKVKLV---EDILAFAPIRHPERLGKPNQAIIYTTDDEIQVAKLAGKYT 529
QY 463 GE-----GVGNIANFELDQPYLPQGTFKYITASKDYDEVSYDGTFTVPTSLAYKMASOTI 517
Db 530 AEDGYIFDPHDITSDEGDVYTPHMTSHHWIKKDSLSE-----AERAAQA- 575
QY 518 FYPTHAGDTYLRVNPQFAVPKGTDALVRVDFPHGNAYLENNYKVGKELKPIPKLNGOTT 577
Db 576 YAXEKLGT-----PPSTDHQDSNGTEAKGAIYNRVKAAK-KVPL----- 615
QY 578 RTAGNKIPVTFMANAYLDNSTYIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKV 637
Db 616 ---DRMPYNEQ-----YTVEV-----KNGSLIIPHYDHYHNKIFEFDEGL 653
QY 638 -EEPTSEKVEKELSETGNSTNSTLEEVPYTPVQEKVAKAESYGMKLENVLFNMDG 696
Db 654 YEAK-----GYTLEDLL-----A 667
QY 697 TIELYL-----PSGEVKKNMADFTGEAPQG-----NGENKPSNGKVGSTGTVENQPTEN 746
Db 668 TVKYVVEHPNERPHSDNGFGNASDHVQRNKGQADTQTEKPSSEKQTEKPEETPRE 727
QY 747 KPADSLPEAPNEKPV-KPENSTDNGLNPEGVSDPMLDSALEEAPVAVPQVQEKLEKFT 805
Db 728 KPQSEKESP--KPTPEEPSEPE-----SEEPQVET-----EKVEEKLEAE 768
QY 806 ASYGLGLDSVI-FNMDGTIELRLPSGEVKKNL 838
Db 769 DLLGKIQDPIKSNAKETL-----TG-LKNLL 795

RESULT 4
Q97QM9 ID Q37QM9 PRELIMINARY; PRT; 819 AA.
AC Q37QM9;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Conserved domain protein.
GN SP1174.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eilen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.F., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,

```

RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,	
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,	
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,	
RT	"Complete genome sequence of a virulent isolate of Streptococcus	
RT	pneumoniae."	
RL	Science 293:498-506(2001).	
DR	EMBL: AE007418; AAK75283.1; -	
DR	TIGR: SP1174; -	
KW	Complete proteome.	
SQ	SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;	
	Query Match 22.5%; Score 989; DB 16; Length 819;	
	Best Local Similarity 31.9%; Pred. No. 1.9e-39;	
	Matches 298; Conservative 118; Mismatches 265; Indels 252; Gaps 36;	
QY	1 CAYALNQHRS-QENKDNRRVSYVDGSGSQSKSENITPDQVSKQEGIOAEQIVIKITDQGY 59	
DB	20 CSYELGRYQAGQDKKESNRVAVIDGQAGKAENLTDPDEVSKREGINAEQIVIKITDQGY 79	
QY	60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNQLKDADIVNEKGGYIKVDGKYVYVLK 119	
DB	80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNQLKDSIVNEIKGGYIKVNGKYVYVLK 139	
QY	120 DAAHADNVRTKEINRQKOEHVKD-NEKXNSVAVARSQGRVTTNDGYVFNPAIDIEDTG 178	
DB	140 DAAHADNIRTKEIKRQKQERSHNSRADNAVAARAQGRITTDGQYIFNASDIIEDTG 199	
QY	179 NAYIVPHGGHYIYIPKSDLSASELAARAKHLAG--NMOPSLSYSTPSP-----SL 229	
DB	200 DAVIVPHGDHYIYIPKNELASASELAARAYWNGKQGRSPSSSSSYNNANPAQPRLSENH 259	
QY	230 PINPGTSH-----EKH-BEDGYGFDAENRTIAEDESGFMVSHGHN 268	
DB	260 TVTTP-TYHQNQGENISLLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGHY 318	
QY	269 HYFFKKDLTEEQIAKQAKLELV-----KTSHNGLDL-----SSHEQDYPSN 311	
DB	319 HF-----IPEQMSLEKRIALIPLYRSNHWVDSRPEEPSPOPTPEPSQPAPSN 373	
QY	312 AKEMKOLDKKIEKTAG--IMQYGVKRVRESIVVNEKKNATIIYPHGDHHAADIDEKHPVG 369	
DB	374 PIDEK-LVKEAVRKVGDVYFEENGYSR-----YIPAKDLSAETAAG 414	
QY	370 IGHSHSNYELFKPEEGVAKKE-----GNKYV-----TGBELTNV 403	
DB	415 IDSKLAKQESLSHLKTKTKTDLPSDRBYNKAYDILLARIHQDLLDNKGRQVDFEALDNL 474	
QY	404 VNLKRNSTFNQNTLIANGQKRVFSFPPELEKKLGINMLVLTDPDGKVLKVGSKVPG 463	
DB	475 LERLKDVSDDKVLV---EDILAFAPIRHPERLKGPNQAITYITDDEIQVAKLAGKYTT 530	
QY	464 E-----GVGNIAINFELDQPLPQOTKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIF 518	
DB	531 EDGYIFDPRDITSDGDAYVTPHMTSHWIKKDSLSE-----AERAAQA--- 575	
QY	519 YPFHAGTYLVRNPOFAVRPGTDALVRVDFEFGNAYLENNYKVGEEKLPKPLNOGTTTR 578	
DB	576 YAKEKGLT-----PSTDHQDSGNTAKGAELYNRKAAR-KVEL----- 615	
QY	579 TAGNKPVTMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKV- 637	
DB	616 ---DRMPYLNQ-----YFVEV-----KNGSLIIPHIDYHNIKFEWFDEGLY 654	
QY	638 EEPKTSKVEKEKLSERGTNSTSLTEEVPTDPOEKVAKFAESYCMKLENVLFNMDGT 697	
DB	655 EAPK-----GYTLEDLL-----AT 668	
QY	698 IEIYL-----PSGEVTKKNMAFTGEAPDG-----NENKPSNGKVSIGTVENQPTENK 747	
DB	669 VKYVVEHPNERPHSDNGFGNASHDVQRNKGQADTNQTEKPSSEKQOTEKPEETPREK 728	
QY	748 PADSLPEAPNEKPV-KPENSTDNGLNPEGNVGDPMDSALEAPADVPQOEKLEKFTA 806	

DB	729 PQSEKPESP-KPTEPEPESPEE-----SEEPQVET-----EKVEKLREARD 769	
QY	807 SYGLGLDSVI-FNMDGTIELRLPSGEVIRKNNL 838	
DB	770 LLGKIQDPIKSNAKETL-----TG-LKNLL 795	
RESULT 5.		
Q9ANY2	ID Q9ANY2 PRELIMINARY: PRT: 839 AA.	
AC	Q9ANY2:	
DT	01-JUN-2001 (Tremblrel. 17, Created)	
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)	
DT	01-OCT-2001 (Tremblrel. 18, Last annotation update)	
DE	Pneumococcal histidine triad protein B precursor (Hypothetical protein	
DE	SP1003) (Fragment).	
GN	PHTD OR SP1003.	
OS	Streptococcus pneumoniae.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;	
OC	Streptococcaceae; Streptococcus.	
OX	NCBI_TaxID=1313;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21101045; PubMed=11159990;	
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,	
RA	Darmatz M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,	
RA	Langermann S., Koenig S., Johnson S.,	
RT	"Identification and characterization of a novel family of pneumococcal	
RT	proteins (the Pnt family) that are protective against sepsis."	
RL	Infect. Immun. 69:949-958(2001).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
PC	STRAIN=TI64;	
PC	MEDLINE=21357209; PubMed=11463916;	
RA	Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,	
RA	Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,	
RA	Urkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,	
RA	Mayday L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,	
RA	Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,	
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,	
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,	
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,	
RT	"Complete genome sequence of a virulent isolate of Streptococcus	
RT	pneumoniae."	
RL	Science 293:498-506(2001).	
DR	EMBL: AF318955; AAK06760.1; -	
DR	EMBL: AE007403; AAK75120.1; -	
DR	TIGR: SP1003; -	
KW	Signal; Hypothetical protein; Complete proteome.	
FT	SIGNAL 29	POTENTIAL.
FT	NON_TER 839 839	
SQ	SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;	
	Query Match 22.5%; Score 987; DB 16; Length 839;	
	Best Local Similarity 32.08; Pred. No. 1.9e-39;	
	Matches 299; Conservative 106; Mismatches 278; Indels 252; Gaps 33;	
QY	1 CAYALNQHRS-QENKDNRRVSYVDGSGSQSKSENITPDQVSKQEGIOAEQIVIKITDQGY 59	
DB	20 CSYELGRYQAGQDKKESNRVAVIDGQAGKAENLTDPDEVSKREGINAEQIVIKITDQGY 79	
QY	60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNQLKDADIVNEKGGYIKVDGKYVYVLK 119	
DB	80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNQLKDSIVNEIKGGYIKVNGKYVYVLK 139	
QY	120 DAAHADNVRTKEINRQKOEHVKDNEKXNSN---VAVARSQGRVTTNDGYVFNPAIDIED 176	
DB	140 DAAHADNIRTKEIKRQKQEH-SHNHGGSGNDQAVAAQAQGRITTDGQYIFNASDIIED 198	
QY	177 TGNAYIVPHGGHYHYIPKSDLSASELAARAKHLAG--NMOPSLSYSTPSP----- 227	
DB	199 TGDYIVPHGDHYHYIPKNELASASELAARAYWNGKQGRSPSSSSSYNNANPAQPRLSENH 258	

[illegible]


```
Db 476 ---ADELAALD-----QEQGKPKLFD-----TKKVS-----RKVTGDKGV 509
QY 454 --LEKVSQVFGVGVGNIANFELDQYLPQGTFF-KYTIASKDYPEVSYDGTFTVPTSLAY 510
Db 510 GYMMPKDGKDY-----FYARDQLDLTQIAFAEQELMLKDKKHRYD---IVDTGTGP 558
QY 511 KMASQTIFFPHAGDTYLRVNPQFAVPKGTQDALRVDFEHGNAYLENNYKVEIKLPIIP 570
Db 559 RLAVDVSSLPFHAGNATYDTGSSFVIPH-----IDHIH----- 591
QY 571 KLNQGTTRTAGNKIPVTMANAYLDNQST--YIVEVP-----ILEKENQTDKPSILPQFK 623
Db 592 -----VVYPSWLTNR---DOIATIKYVMQHPVEVRPDINSKPGHEESGVIPNV- 635
QY 624 RNKAQENSKLDEKVEEPK-----TSEKVEKEKLSGTNSTSLTELEVPDVPQVKVAK 678
Db 636 -----TPDKRAGMPNMQIIHSAEEVQK-ALAEGRPATPDGYI-----FDP-RDVLA 681
QY 679 FAESYGMKLENVLFNMDGTIELYLPSEGVIKK-NWADFTGEAPQNGENKPSGVSTG 737
Db 682 ---ETFWVK-----DGSPISPRADGSSLRITINKSDLSQAEWQAQELLAKKNAGDAT 731
QY 738 TVENQPTNKPADSLPE--APNEKPVKPNSTDNGLNPNPEGNVGSDPMLDSALEEPAVD 795
Db 732 T--DKPKQQAQKSNENQPSAEKEEKESDDFI-----DSLDPYGLDRATLED 780
QY 796 PQVEKLEKTASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 781 HINLAQK-----ANIDPKYLIFQPEG-VQFYKNGELV 813

RESULT 9
Q99XV4
ID Q99XV4 PRELIMINARY; PRT: 825 AA.
AC Q99XV4:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY2006.
GN SPY2006
OS Streptococcus pyogenes
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=13114;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11298296;
RA Ferrer J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006623; AAK34688.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF50969E650F4 CRC64;

Query Match 19.7%; Score 864; DB 16; Length 825;
Best Local Similarity 28.4%; Pred. No. 1, 4e-33;
Matches 267; Conservative 144; Mismatches 273; Indels 256; Gaps 43:

QY 2 AYALNQHRSQENKNNRVSVDGSSQSKS--ENLTPQVQKEQIAEQIVIKITDQGY 59
Db 22 SYQLGKHMGMSATKDNQAIYIDDSKGAKAPKTKMTDQISAEGISAEQIVWKITDQGY 81
QY 60 VTSHGDDHYHYNGKVPYDALFSEELLMDPNVQLXADIVNEKGGYLIKDGKYYVYLK 119
Db 82 VTSHGDDHYHYNGKVPYDALFSEELLMDPNVQLXADIVNEKGGYLIKDGKYYVYLK 141
QY 120 DAAHADNVRTKDEINRQKQEHVKD-NEKVNNSVA-----VARSQGRYTINDGYV 167
Db 142 PGSKRNRTKQIAEQVAKGTFKEAKERGLAQVLAHLSKEEVAANAEKRGQRYTIDGYI 201
```

```
QY 168 FNPADLIETGNTAYPHGHYHYIPKSDLSASELAARAKHLA---GKNMOPSOISYST 224
Db 202 FSPDTDIIDDLGDAYLVPNGNHYHIIPKDKLSPELAARAAQAIYWKQGRGARPS---DIRPT 259
QY 225 PSPS-----LPT-----NPGTSHE-----KH----- 240
Db 260 PAPAPGRRKAPIPDVTNPNPGOGHGDNGGYHPAPPRPNDAQNKHQRFDEFKGKTFKELLD 319
QY 241 -----EEDGICFDANRIIAEDSCFVMSHGDHNNHYFFKDKLTERQIYAAQAKHL 288
Db 320 QLHRLDLKYRVEDGLIFEPTQVIKSNAGFYVYPHGDHYHIIIPRSQLSPELEMLADRY- 378
QY 289 BEVKTSHNGDLSHSHEQDYPSNAKEMKDKDKTEEKIAG-IMQYGV-----KRES 339
Db 379 -----LAGQTEDDSGSDHSPKSPDKVTHFLGHRIRKAYKGLDKGKYDTSDA 426
QY 340 IVVNKE-----KNAILYPHGDHHAADPIDHKPVGVGCHSHSNYELFKPESGVAKKGNK 393
Db 427 YVFSKESITHSVKSGVTAHGDHFHY-----IGFEL-EQYELDEVANWV-KARQV- 475
QY 394 VYTGEELTNVNLKNSFTNQNTLANGOKRVSFSPPELEKKGILGNMLVKLITPDGKV 453
Db 476 ---ADELAALD-----QEQGKPKLFD-----TKKVS-----RKVTGDKGV 509
QY 454 --LEKVSQVFGVGVGNIANFELDQYLPQGTFF-KYTIASKDYPEVSYDGTFTVPTSLAY 510
Db 510 GYMMPKDGKDY-----FYARDQLDLTQIAFAEQELMLKDKKHRYD---IVDTGTGP 558
QY 511 KMASQTIFFPHAGDTYLRVNPQFAVPKGTQDALRVDFEHGNAYLENNYKVEIKLPIIP 570
Db 559 RLAVDVSSLPFHAGNATYDTGSSFVIPH-----IDHIH----- 591
QY 571 KLNQGTTRTAGNKIPVTMANAYLDNQST--YIVEVP-----ILEKENQTDKPSILPQFK 623
Db 592 -----VVYPSWLTNR---DOIATIKYVMQHPVEVRPDINSKPGHEESGVIPNV- 635
QY 624 RNKAQENSKLDEKVEEPK-----TSEKVEKEKLSGTNSTSLTELEVPDVPQVKVAK 678
Db 636 -----TPDKRAGMPNMQIIHSAEEVQK-ALAEGRPATPDGYI-----FDP-RDVLA 681
QY 679 FAESYGMKLENVLFNMDGTIELYLPSEGVIKK-NWADFTGEAPQNGENKPSGVSTG 737
Db 682 ---ETFWVK-----DGSPISPRADGSSLRITINKSDLSQAEWQAQELLAKKNAGDAT 731
QY 738 TVENQPTNKPADSLPE--APNEKPVKPNSTDNGLNPNPEGNVGSDPMLDSALEEPAVD 795
Db 732 T--DKPKQQAQKSNENQPSAEKEEKESDDFI-----DSLDPYGLDRATLED 780
QY 796 PQVEKLEKTASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 781 HINLAQK-----ANIDPKYLIFQPEG-VQFYKNGELV 813

RESULT 10
Q92HG7
ID Q92HG7 PRELIMINARY; PRT: 822 AA.
AC Q92HG7:
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R3268;
RX MEDLINE=99115568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schnitzler N., Luetticken R., Podbielski A.;
RT "Lmb, a protein with similarities to the Lrai adhesin family, mediates
```

RT attachment of Streptococcus agalactiae to human laminin.";

RL Infect. Immun. 67:871-878(1999).

DR EMBL; AF062533; AAD13797.1; .

KW Hypothetical protein.

SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

```

Query Match          19.5%; Score 855.5; DB 2; Length 822;
Best Local Similarity 29.0%; Pred. No. 3.6e-33;
Matches 270; Conservative 143; Mismatches 277; Indels 241; Gaps 45;

QY 2 AYALNQRHSQENKDNRRVYVDSQSQSKS--ENLTPDQVSQKEGIAEQIIVIKITDQGY 59
DB 22 SYOLGKHMGKATKDNQIAYIDDSKGVKAPKTKNTMDQISAEEGISAEQIIVKITDQGY 81
QY 60 VTSCHDHYHYNGKVPYDALFSEELLMKDPNVLKADIVNEVKGGYIIVKDGKYYVYLK 119
DB 82 VTSCHDHYHYNGKVPYDAIISELLMTDPNHFQKQSDVINEILDGVIIVKNGNYYVYLK 141
QY 120 DAHADNVRTKDEINRQKQEHVKD-NEKVNNSVA-----VARSGRYTTNDGYV 167
DB 142 PGSKRNIKTQKQIAEQVAKGTKEAKEGLAQVAHLSKEEVAANVEAKRGQRYTTDDGYI 201
QY 168 FNPADIIEDTGNAYIVPHGHHYHYIPKSDLSASELAHAAHIA--GKNMQPSQLSYST 224
DB 202 FSPTDIIDLDGDAYLVPHGNHYHYIPKDLSPSELAAQAYWSQKQGRARPS--DYRPT 259
QY 225 PSP---SLPT-----NPGTSHE-----KH----- 240
DB 260 PAPGRKAPIDVTPNPGQGHQDPNGGYHPAPPNPNDASQNHQDEFFKGTFFKELLDL 319
QY 241 -----EEDGYFDANRITAEDESFGVMSHGHDHNYFFKDLTEQIKAAQKHLEE 290
DB 320 HRLDLKRYHVEEDGLTFEPTQVTKSNAGFYVPHGHHYHYIPRQSLSPLEMLADRYLAG 379
QY 291 VKTSHNGCLDSLSEHQDYPSNAK-----ENKLDLKIEEKIAGIMQYGVKRESIVV 342
DB 380 -QDDN--DSGSDHSK--PSDKVETHTF.LGRIKAYKGLDQKPYDTSAYVFSKESI-H 433
QY 343 NKEKNALIVPHGHHADPIDEHKPVGIGHSHSNYELFKPEGVAKKEGKNKYVTGELTN 402
DB 434 SVDKSGVATKEGDFHY-----IGFGL-EGYELDEVANVW-KAKGO-----ADELVA 479
QY 403 VVLLKNSNTENNQNFILANGOKRVYSFPPELEKLGILNVLKLTIPDGKV--LEKVSCK 460
DB 480 ALD-----QEOGKERPLED-----TKKVS-----RKVTGDKGVYIMPKDGR 516
QY 461 VFEGGVGNIANFELDQPYLPQGFQKTYIAASKDYPEVSYDGTFTVPTSLAYKVASQTI 520
DB 517 DY-----FYARYQLDLTQI--AFARQELMLKDKKHRYD--IVDGTIEPLRAVDYSSLP 566
QY 521 FHAGDTYLRVNPQFVAPKGTDALRVDFEFGNAY--LENNYKVGIEIKLPIPKLNOGTR 578
DB 567 MHAGNATYDGSFSFVIPH-----IDHIVVYPSWLTNR-OIATIK----- 605
QY 579 TAGNKIPVTMANAYLDNQSTYIIVEP-----ILEXENOTDKPSILPQFKRNKAQENSKL 633
DB 606 -----YVQHPREVDPVWSRPGHGESGVIPNV-----TFL 636
QY 634 DEKVEEPK-----TSEKVEKEKLESETGNSTNSLTLEVPVDPQVKVAFESYGMKLE 688
DB 637 DKRAGMNPWQIIRSAEEVQK-ALAEGRFAAPDGYI-----FDP-RDVLAK--ETFWVK-- 685
QY 689 NVLFNMDGTTIELYLPSEVILK-NMADFTGEAPQNGENKPSNGKRVSTCTVENQPTENK 747
DB 686 -----DGSFSIPRADGSSURTINKSDLSAQEQWQAEQELLAKKNAGDATDT--DKPEEK 737
QY 748 PADSLPAPNEKP---VKPNSNDSTNGMLNPEGNVGDPMPLDSALEAPAVDPVQEKLEKF 804
DB 738 QADKSNE--NQQPSSEASKEEKESDDFI-----DSLTPDYGLDRATLEDHINOLAQK- 785
QY 805 TASYGLGLDS--VIFNMKGHIELRLPSGEVI 833
DB 786 -----ANIDPKYLIFQPEG-VQFYNKNGELV 810

```

```

RESULT 11
Q9AE21
ID Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBSII, an active group
  II intron identified in human isolates of group b streptococci.";
PL J. Bacteriol. 183:2560-2569(2001).
DR EMBL; AJ290952; CAC35985.1; .
KW Hypothetical protein.
FT NON_TER 289 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;

Query Match          15.7%; Score 689; DB 2; Length 289;
Best Local Similarity 51.3%; Pred. No. 8.3e-26;
Matches 138; Conservative 42; Mismatches 61; Indels 28; Gaps 8;

QY 2 AYALNQRHSQENKDNRRVYVDSQSQSKS--ENLTPDQVSQKEGIAEQIIVIKITDQGY 59
DB 22 SYOLGKHMGKATKDNQIAYIDDSKGVKAPKTKNTMDQISAEEGISAEQIIVKITDQGY 81
QY 60 VTSCHDHYHYNGKVPYDALFSEELLMKDPNVLKADIVNEVKGGYIIVKDGKYYVYLK 119
DB 82 VTSCHDHYHYNGKVPYDAIISELLMTDPNHFQKQSDVINEILDGVIIVKNGNYYVYLK 141
QY 120 DAHADNVRTKDEINRQKQEHVKD-NEKVNNSVA-----VARSGRYTTNDGYV 167
DB 142 PGSKRNIKTQKQIAEQVAKGTKEAKEGLAQVAHLSKEEVAANVEAKRGQRYTTDDGYI 201
QY 168 FNPADIIEDTGNAYIVPHGHHYHYIPKSDLSASELAHAAHIA--GKNMQPSQLSYST 224
DB 202 FSPTDIIDLDGDAYLVPHGNHYHYIPKDLSPSELAAQAYWSQKQGRARPS--DYRPT 259
QY 225 PSP---SLPT-----NPGTSHEKHEEDGY 245
DB 260 PAPGRKAPIDVTPNPGQGHQHP-DNGGY 287

RESULT 12
Q99Z76
ID Q99Z76 PRELIMINARY; PRT; 792 AA.
AC Q99Z76;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative internalin A precursor.
GN INLA OR SPY1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
  Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
  Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
  Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

```

```

RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: A606574; AAK34188.1;
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_Out.
DR Pfam: PF00560; LRR; 4.
DR SMART: SM00370; LRR; 4.
DR Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32288485ACE0 CRC64;

Query Match
Best Local Similarity 5.1%; Score 224.5; DB 16; Length 792;
Matches 175; Conservative 114; Mismatches 279; Indels 333; Gaps 43;

QY 3 YALNHRSEKNDNRFVSDGSQSQSKSEN-----LTPDVSQKEGIAE 48
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
31 YPKTKQSRGKMTSNKIKPKSKTKTKTKGVAGVDFPTDDGFLTKDSK----- 81
QY 49 QIVIKITDQGYTSHGDHYHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNKGVI 108
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
82 --ILSKTDQGVVVDHGHSHF-----IFYADL----- 106

QY 109 KYDGKYYVYL--KDAHADNVTKDEINRQKQEHVKDKNVSNVAVARSQGRYTND-- 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
107 --KGSFXYLLPKGASLA-----KPAVQR-----AASQTSKVADPH 142

QY 165 -GYVFNPAIDIEDTGNAYIVPHGGHYHYPKGDLASSELAAAKAHLAGKNMPSOLSYSS 223
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
143 HHYFNPAIDIAEDALGYIVRHDDHEHYLLKSLSCQTQAAK-QVAFRLPOTSLVSTA 201

QY 224 TSPSLPINTGSHKEHEEDGGVDFANRIADESGFVNSHGHDHNYFFKXDLTE----- 278
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
202 TANGITPLGFPTS-----DGFQNGQGVGVTKDSILVDHGHLPISFADLRQGGWAH 255

QY 279 -----FOIKAAQKHLEVKVTSNGLDLSLSSHEQDYPYNAKENKDKDKTEEKIAGIMQY 333
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
256 VADQYDPAAKAKPAETHQT-----PELSEREKEVQEKLATLAEKL 296

QY 334 GYKRSI-----VYNNKKNALIVPHGDH-----HRAFDIDHKFVG 369
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
297 GIDPSTIKRVTQDGKLGLEYPHDHAHVHLSLDIEIGKIDPIDPAIEHARELEKH-KVG 355

QY 370 -----IGHSHSNYELFPEEGVAKKEGKVTGTGELTNVNL---LNKS 410
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
356 MOTLRALGPDEVIDIVKTHDAPTFPSNE-----KDNMM--KEWLATVTKLDLGRSKD 409

QY 411 TFNNQNTLANGOKRVSFSPPELE-----KLGINMLVKLIIPDGKLVKSVKVF 463
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
410 PLQRKGLSLPLNLETIGIGFTPIKDIPVLPQFKLKLQLLMKTGYTDYRFLD----- 461

QY 464 EGVGNIAFELDQVLPQGTFTYATSKDYPVSDGTFTVPTSLAYKMASGTIPYFPHA 523
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
462 -----NMPLEG-----IDISQNLKDISFLSKYKNTLVA-----A 493

QY 524 GDTYLR-VNQPFAVPGKTDALVRVEDEFHGNAYLENNYKVGEIKPIPKLNOGTTTAGN 582
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
494 ADNGIEDIRLQGLP-NLKFLV-----LSNN-KISDLS-PLASLHQ----- 531

QY 583 KIPVTFMAYLDNOSTVIVEV-PILEKENQT-----DKPSI----- 618
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
532 -----LOELHIDNQ--ITDLSPSVSHKESLTVDLSRNADVLDLATLOAPKLETLMWNT 583

QY 619 -----LPQEKRNAQESKLDEKVEEPTSEKVEKEKELSTGNSTNSTLEE 665
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
584 KYSHLDLFKNPNLSLSINRAQLSL--EGIEASSVIVRVEAE-----GQIKSLVLKD 636

QY 666 VPTVDPQVEKVAFAESYGMKLENLYFNMDG-----TIELYLPGEV 707
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
637 -----KQSLTFLDVTG---NQLTSLEGVNNFTALDILSVSNQLTNVNLSPKNTV 685

QY 708 -----IKKN---MADF--TGEAPOGNGENKPS--ENGKVSTGTVENQOT-ENKPADSLPE 754
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
686 TNIDISHNNSIADLKLEQHPIPEIAKFNFAVYEGSMVNGTAEKAAATKAKESAQE 745

```

```

QY 755 A 755
Db 746 A 746

RESULT 13
Q99RD3
ID Q99RD3 PRELIMINARY; PRT: 961 AA.
AC Q99RD3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE fNB protein (Fibronectin-binding protein homolog).
GN fNB OR SAV2502 OR SA2290.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
Staphylococcus.
NCBI_TaxID=158878, 158879;
[1]
SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; Pubmed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani J.-I., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003365; BAB58664.1;
DR EMBL: AP003367; BAB43593.1;
DR InterPro: IPR004237; Fb_bind.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF02986; Fb_bind; 1.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR01217; PRICHTEXTENS.
DR TIGRFAMs: TIGR01167; LPXTC_anchor; 1.
DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 961 AA; 106010 MW; 364940F884EA4101 CRC64;

```

```

Query Match
Best Local Similarity 4.8%; Score 212.5; DB 16; Length 961;
Matches 176; Conservative 110; Mismatches 287; Indels 251; Gaps 42;

QY 26 QSSQKSE-NLTPDVSQKEGIAEQIVIKITDQGYTSHG--DRHYHYNGKVPYDALFSE 82
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
290 QGSQKVEVTLQCNKVSKEFDI---KYLDGVKDRMGVTVNGRIDTLNKEEGK-----FSH 340

QY 83 ELLMKDPNQLKADIVNEVKGVIIVKVG---KYVYLKDAHADNVTKDEINRQKE 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
341 FAYVKNPNQSLTSVTVTQVTSQVYKQANNPVTVKYKHHGSDLAESVYAKLD-DTSKFE 399

QY 140 HVKNEKVNVAVARSGRYTTNDGVYKPADLIETGNAYIVPHGGHYHYTPKSDLSA 199
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
400 DV--TEKNLS-----YTSNGGYTLNLGDL--DNSKDYIVIKYEYGEY-----DQNA 440

QY 200 SELAAKAHLAGKNMPSQLSYSTSPSPSPINPCTSHEKHEEDGYGFA--NRITAEDE 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
441 KDL-NFRTHLSGYHKYYPYYPYYPVQLTWNNGVAFYSNNAKGDKGKPNPIEKSE 499

QY 258 SGFVMSGHDHNYFFKDLTEEQKAAQKH-----LEEVTSS-----HNGDLSLSSE 305
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
500 -----PIDLDIKSEPPVEKHELTGTIEESNDSKPIDFEYHTAVEGAEGHA 544

QY 306 QDYPSTNAKEMKDLDKTEEKIAGIMQYGVKRESIVVYNNKKNALIVPHGDHHDHPIDEH 365

```

Db 545 E-----GIIIE-----TEEDSIHDFEEST---HENSXKHADVVEYE 577
Qy 366 KPVGIG-----HSHSNYLFKPE-EGV-----AKKGNKYVTGEELTNVNL-----406
Db 578 EDTNPGGGQVVTESNLVDFEESTKGLVTCVSDHITVEDTKYITE--SNLIELVDLP 635
Qy 407 -----LNKNFTNONT-----LANGQKRVSFSPFPELEKGLGIMLVKLITPDGKYL 454
Db 636 BEHGQAQPIBEITENNHHISHGLGTENGHNGYGVIDEIENSHVDIKSEL-----687
Qy 455 EKVSQVFGEGVGNIAFNDQPLPQGTFTYIAKSDKPEVSDGTFTVPTSLAYKMAS 514
Db 698 -----GYEGQNSGN-QSFEED-----TEDRKPYEQGNI-----717
Qy 515 QTIFYPFHAGDTYLRVNPQFAVPGKTDALRVDFEFGHAYLNNYKVGELKIPKLNQ 574
Db 718 -----VDIDF-----DSVPQIHQNGNGQSFEDTEEDK-----PKYEQ 751
Qy 575 GTTRTAGNKIPVTMANAYLDNSTYIVVPILKEKNOTDKPSILPQFKRKAQENSKLD 634
Db 752 G-----GNLIDIDFSDVPQIHGNKH-----NEIIEEDTKDKPNY---QF-----GGHNSYDF 797
Qy 635 EKVEEPTSEKVEREKLSETGNTSNTLSEVPT-----VDPVQEKVAKFAESYGMKLEN 689
Db 798 EEDTLPKVSGQNEQQTIEEDTTPPTPTPEVPSEPTPTPTPTPEVPSPGEPPTPKPE- 856
Qy 690 VLFNMDGTIELYLSGVEIKKNADTFGEAPQNGE-----NKPSNGKVSSTG 737
Db 857 -----VPSEPTPEVP-----TPVPSGPKVPVPAKEEPKPKPKVQGGKVVTP 901
Qy 738 TVE-NQ-----PTENKPA--DSLPEAPNEKPKVKNSTNDGML 772
Db 902 VIEINERKAVAPTQKQSKKSELPETGEE-----STNKGML 939
RESULT 14
ID Q97Q7 PRELIMINARY; PRT: 2004 AA.
AC Q97Q7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
GN SP1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TI6R4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anghuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EXBL: AE007416; AAK75263.1;
DR MEROPS: M26.001;
DR TIGR: SP1154;
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001130; Zn_MTPetdse.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004_AA: 223908 MW: 556BC6A1028D60AA CRC64;
Query Match 4.8%; Score 210.5; DB 16; Length 2004;
Best Local Similarity 22.1%; Pred. No. 0.098;
Matches 176; Conservative 93; Mismatches 241; Indels 295; Gaps 47;
Qy 186 GGHYHYPKSDLSASELAALAAHACKNMOPSLSY-----SSPP 225
Db 47 GVHYKYVADSLSEE-----KKQLVYDIPTYVENDDETLYLVKLNQ 92
Qy 226 SPSLPINPGTSHKHE--EDGYGFDAIRIAEDSGVMSHGDNHFFKDKTEQIRAA 284
Db 93 LAELP-NTGSKNEHQALVAGASLAAMGILI-----FAVS-----KKVKNKT 134
Qy 285 QKHEERYKTHNG-----LDSLSSH-----BDYPSNAKEMKDLKRIE-EKIAGINKQY 334
Db 135 -LHLVLVAGTNGVLSVSHALENHLNLYNTDYELTSGEKLPLPKETISGYTVGYKEGK 193
Qy 335 VKRSIVVNEKNAIIVP-----HGDH-----HHADPIDERKPVGIG--- 371
Db 194 TTSSEVSN-QKSSVATPTKQKVYVNTFVDFHSTVQAIQEQTPVSTKPTVEQVVE 252
Qy 372 -----HSHSNYELFKPEGVAKREGKRYT-----GELTNV 403
Db 253 KPFSTELINPRKEKQSSDSQEOIAEHKNET-KKEKISPKETGVNTLNQDEVLSQ 311
Qy 404 VN-----LLKNSTFN-----NQFTLANGOKRVSFSPFPELEKGLINM-LYKLIT- 448
Db 312 LNKPELLYREETMETKIDFQERTQENPDLAEGTVRV-----KQEGKLGKVEIVRIEVS 365
Qy 449 -----PDGKVLKVSCK--VFGE-----GV-----GNIAFELDOPYL 479
Db 366 NKEEVSREIVSTSTAPSPRIVEKTKTKTQVKEQETGVHEKDKVQSGALVEPAL-QPEL 424
Qy 480 PGQTFKYTIASKDYPEVSYDGTFTVTSLAYKMASQTIYPFHAGDTYLRVNPQ----FA 535
Db 425 P-----EAVSDKGEPEVQ-----TLPEAVV-----TDKGET--EVQPESDTVV 463
Qy 536 VPKGTALRVDFEFGHAYLNNYKVGELK--LPKPL-NQOFTTAGNKIPVTFMANA 592
Db 464 SDKGEPEQVAPLPEYKGN-----IEQVKPETVKTKEQGEPT--EEVpV----- 507
Qy 593 YLDNQSTYIIVEVPILKEKNOTDKPSILPQFKRN---KAOENSKLDEKVEEPTSEK--- 645
Db 508 -----KPTETPNPNEGTEGTSI--QDAENPVQPAEESTTNSSEK-SPOTSSKNWGE 558
Qy 646 -----VEKEKLSGTGNTSNTLSEVPTVDPVQEKVAKFAESYGMKLENYL 691
Db 559 VSSNPSTSTSVGESNKNPEHNSKENSEKTEVEVP-VNP----- 597
Qy 692 FNMGTIELYLPSCVEIKKNADTFGEAPQNGENKPESENGKV---STGTVENQPTENKP 748
Db 598 --NEGTEV-----GTSNQETEKVPQPAEETQTSNGKIANENTGEVSNKPSDSKP 644
Qy 749 --ADSLPEAPNEKPKVPKENS-----TDNGMLNPE---GNVGSDFMLDSALEAPVDPVQE 799
Db 645 PVESNQPEKNGATKATPEPSNGTNSGNTSENGOTEPPEPSNGNSTEDVSTESNTSNGNEELIQ 704
Qy 800 KLEKFTASYGLGLDSVIFNMDGIELR 826
Db 705 ENE-----LDPPDKVVEEPEKTELEUR 724
RESULT 15
ID Q963T1 PRELIMINARY; PRT: 1078 AA.
AC Q963T1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:52 ; Search time 13.0205 seconds
(without alignments)
1898.176 Million cell updates/sec

Title: US-09-471-255-16
Perfect score: 4396
Sequence: 1 CAVALNHRSQENKNNRVS.....GTIELRPSGEVKKNLLIS 840

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1337.5	30.4	447	4	US-08-961-083-182
2	980	22.3	763	4	US-08-961-083-66
3	966.5	22.0	796	4	US-08-961-083-56
4	197	4.5	571	4	US-08-961-083-4
5	194.5	4.4	1964	2	US-08-790-912-3
6	194.5	4.4	2052	2	US-08-790-912-2
7	181.5	4.1	10182	4	US-09-134-001C-3159
8	180.5	4.1	984	1	US-08-242-932-2
9	180.5	4.1	984	1	US-08-714-481-2
10	180.5	4.1	984	5	PCT-US95-06111-2
11	180.5	4.1	1848	4	US-08-296-791-6
12	180.5	4.1	1848	5	PCT-US95-10661A-6
13	180	4.1	1231	3	US-08-904-263A-4
14	180	4.1	1435	2	US-08-568-459A-4
15	180	4.1	1435	2	US-08-487-826B-4
16	180	4.1	1435	4	US-09-210-288-4
17	178	4.0	1104	4	US-08-923-992A-4
18	175.5	4.0	1164	4	US-08-923-992A-2
19	174	4.0	1098	4	US-08-923-992A-8
20	173.5	3.9	1128	4	US-08-923-992A-6
21	173	3.9	1183	2	US-08-447-031A-2
22	172	3.9	1164	4	US-08-923-992A-10
23	166	3.8	962	4	US-09-071-035-246
24	166	3.8	962	4	US-09-071-035-250
25	166	3.8	962	4	US-09-071-035-254
26	166	3.8	962	4	US-09-071-035-470
27	166	3.8	962	4	US-09-071-035-474

28	166	3.8	962	4	US-09-071-035-478	Sequence 478, App
29	165	3.8	1541	4	US-08-296-791-3	Sequence 3, Appl1
30	165	3.8	1541	5	PCT-US95-10661A-3	Sequence 3, Appl1
31	164	3.7	1545	4	US-08-296-791-4	Sequence 4, Appl1
32	164	3.7	1545	5	PCT-US95-10661A-4	Sequence 4, Appl1
33	162	3.7	1702	4	US-08-296-791-5	Sequence 5, Appl1
34	162	3.7	1702	5	PCT-US95-10661A-5	Sequence 5, Appl1
35	157.5	3.6	894	4	US-09-071-035-248	Sequence 248, App
36	156	3.5	278	4	US-08-961-083-94	Sequence 94, Appl
37	153	3.5	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
38	152.5	3.5	1164	2	US-08-589-756-1	Sequence 1, Appl1
39	152.5	3.5	1164	4	US-09-206-800-1	Sequence 1, Appl1
40	152.5	3.5	1164	4	US-09-206-898-1	Sequence 1, Appl1
41	151.5	3.4	851	4	US-09-071-035-326	Sequence 326, App
42	151.5	3.4	851	4	US-09-071-035-330	Sequence 330, App
43	151.5	3.4	851	4	US-09-071-035-334	Sequence 334, App
44	151	3.4	783	6	5231168-2	Patent No. 5231168
45	151	3.4	2032	4	US-09-071-035-458	Sequence 458, App

ALIGNMENTS

RESULT 1
US-08-961-083-182
Sequence.182, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 30.4%; Score 1337.5; DB 4; Length 447;
Best Local Similarity 60.4%; Pred. No. 4.5e-93;
Matches 270; Conservative 0; Mismatches 0; Indels 177; Gaps 1;

OY 5 LNQRHRSQENKNNRVSVDGSSQSKSENLTDPQVSQKEIGIAEQIVIKITDQGYVTSHG 64
|||||
DB 1 LNQRHRSQENKNNRVSVDGSSQSKSENLTDPQVSQKEIGIAEQIVIKITDQGYVTSHG 60

QY 65 DRYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYYIKVDGKYVYVYLKDAHA 124
DB 61 DRYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYYIKVDGKYVYVYLKDAHA 120
QY 125 DNVRTKDEINROKQEHVNDKNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGNAYIVP 184
DB 121 DNVRTKDEINROKQEHVNDKNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGNAYIVP 180
QY 185 HGGHYHYTPKSDLSASELAALAAKAGLAGNMOPQSLSYSTASDNTQSAKGSKPANK 240
DB 181 HGGHYHYTPKSDLSASELAALAAKAGLAGNMOPQSLSYSTASDNTQSAKGSKPANK 240
QY 224 ----- 223
DB 241 SENLQSLKELYDSPAQRYSSEGLVDFPAKIISRTNGVAIPHGDHYHIFPYKLSAL 300
QY 224 ----- 223
DB 301 EKIARMPVISTGTSTVNTAKPNEVWSSLSLSPNPSLTSKELSSASDGYIFNPKDI 360
QY 224 -----TPSPSLPINPGTSHEKHEEDGYGF 247
DB 361 VEETATAYIVRHGDHFHYPKSNQIQPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGF 420
QY 248 DANRIAEDESGFVMSHGDHNYFFKK 274
DB 421 DANRIAEDESGFVMSHGDHNYFFKK 447

RESULT 2

US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match 22.3%; Score 980; DB 4; Length 763;
Best Local Similarity 31.0%; Pred. No. 1e-65;
Matches 266; Conservative 123; Mismatches 223; Indels 292; Gaps 39;
QY 1 CAVALLQHRH-QENKONNRVSYVDGSSQSKSENLPDQVQSKEGTOAEQIVTKITDQGY 59
DB 1 CSVELGRHQAGQVKESNRVSYIDGQAGQKAENLPDDEVSKREGINAEQXVIXITDQGY 60
QY 60 VTSRGDHYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYYIKVDGKYVYVYLK 119
DB 61 VTSRGDHYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYYIKVDGKYVYVYLK 120
QY 120 DAHADNVRTKDEINROKQEHVNDKNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGF 178
DB 121 DAHADNVRTKDEINROKQEHVNDKNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGF 180
QY 179 NAVTVPHGGHYHYTPKSDLSASELAALAAKAGLAGNMOPQSLSYSTSP-----SL 229
DB 181 NAVTVPHGGHYHYTPKSDLSASELAALAAKAGLAGNMOPQSLSYSTSP-----SL 229
QY 230 PINPGTSH-----EKH-BEDGYGFDANRIAEDESGFVMSHGDGN 268
DB 241 TVTP-TVHQOGENISLLRELYAKPLSERHVESDGLIEDPAQITRTARGVAVPHGNHY 299
QY 269 HYFFKDLTEEQIAKAKHLEEV-----KTSNGLSLSHSEODYPSNAKEMKDLDKKE 323
DB 300 HF-----IPYEQMSELEKRIARIIPLYRSNHWVPD--SRPEQSPQSTPEPSPQPA- 351
QY 324 EKTAGIMKQYGVKRESIVNKEKNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPE 383
DB 352 -----PNQPAPSNPIDE----- 364
QY 384 EGVAKKGNKVVYTGEBELTVNVNLLKSTFNNQNTLANGOKRVSPFPPELEKKGILNML 443
DB 365 -----KLVK-----EAVRVKG----- 375
QY 444 VKLITPDGKYLEKVGKVGEGYGNIANFELDQYLPQGTFKYTIASKDYPEVSDGTFT 503
DB 376 -----DGYVFE-----NGVSR-----YIPAKDLSAETAA-----GIDSKLA 407
QY 504 VPTSLAYKMASQIFYPFHAGDTYLRVNPQFAYPKGTDALVRVDFEHGNAYLENNYKV- 562
DB 408 QESLSHKLGAKKTDLPSSDREFY-----NKAYDLLARIHQDLLNKGQVDFEAL 458
QY 563 -----GEIKL-----PI-----PKLNQTTTETAGNKIPVTFMANAYL- 594
DB 459 DNLLERLKVXSKVKLVXDPILAPIRHPERLGRPNQAQITVT-DDEIQVAKLAKYTT 517
QY 595 -----DNQSTYIVVEPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEERK 641
DB 518 EDGYIFDPRDITSDGDAYV--TPHMTSHWIKKDS--LSEAEAAAQAYAK-EKGLTPPS 573
QY 642 TSEKVEKEKLSGTNSNSTLEEVPTDPVQEKVAKFAESYGMKLENVLFNMDGTIE-- 699
DB 574 TDHQ-----DSGNT-----EAKGAEAIYNRV-KAAKK--VPLDRMPYNLQIYIVVK 616
QY 700 -----LY-LPSEVTK-----KNMADTGEAPQ-----GNG 724
DB 617 NGLIIPHYDHYHNKPEWDFEGLYEAPKGYTLEDELIATVYVVEHPNRPKPHSDNGFGNA 676
QY 725 EN--KPSENGKVGSTGVENOPTENKPADSLPEAPNKPVKPENSTONGMLNPGNVGSDP 782
DB 677 SDHVORKNQOATNQTE-KPSEKPKQTERPEBETPREKQSE-----KPSPKPTTE 729
QY 783 MLDSALE-EAPAV--DPVQEKLEK 803
DB 730 PEESPESEEPQVETEKVEEKLRE 753
RESULT 3
US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469

175 EDGNAYIVPHGHYHYIPK-SDLASASELAALAHAGNMQPSQLSYSTSPSLPIMP 233
Db 19 EISGYTYI-----GYIREKGTISESVNQSSVATPKQ-QKDYINVT--PNFVDHP 68
Qy 234 GTSHERHEEDGYGFDANRIIAEDSGFVMSHGDHNYFFKDLTBSQIAAKQHLEEVKT 293
Db 69 ST-----VOAIEQTPVSTKTEVQVVEKFFSTELINPR-----KEEQ 108
Qy 294 SHGLDLSLSEHQDPSNAKEMKDLKIEKIAGIMKOYGVAKRESIVVYKKNALIIYPH 353
Db 109 SSDSQEQLAEH-----KNLETKKEEKIS-----PKEKTGV----- 138
Qy 354 GDHHDADPDE-----HKPVGICHSHSNVYELFKPEGVAKREGNKVYVGEELTNVNL 407
Db 139 ---NTLPQDEVLSGQLNKP-----ELLYREETE-----TKIDFQEEI----- 174
Qy 408 KNSTFNQNTFLANGOKRVSFSPPELEKLGIMN-LVKLIT-----P 449
Db 175 -----QENPDLAEGTVRV-----KQEGKLGKKVEIVRIFSVMKEEVSREIVSTSTTAP 222
Qy 450 DGKVLKVSCK--VFGE-----GV-----GNANFELDQPYLPQGTFFKTTASKDYPEV 496
Db 223 SPRIVERGTRKTQVIREQPETGVHEKDVQSGAIVEPAI-QPELP-----EAVVSDKGEPEV 277
Qy 497 SYDGTFTVPTSLAYKMASQTIFFPHAGOTYLRVNPQ-----FAVPKGTDALVRVDFPHG 552
Db 278 QP-----TLPEAVV-----TDKGET--EVQPESPDTVVSDKGEPEQVAPLPEYKG 320
Qy 553 NAYLENNKYVGEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSTIYIVVEVILEK 609
Db 321 N-----IEQVKPETPVEKTEQGEPTK--BEVPV-----KPTETPVNPN 358
Qy 610 ENOTDKPSILPQKRN---KAQENSKLDERKVEPKTSEK-----VEK 648
Db 359 ECTTEGTSI--QEAENPVQAEESTINSEKV-SPTDSSKNTGEVSNPDSSTTSVGSNK 415
Qy 649 EKLSETGNTSNTSLVEEPTVDPVQEVKVAFAESYGNKLENLVFNMDGTTIELYLPGEVI 708
Db 416 PEHNDKSNSEKTEVEVP-VNP-----NEGIVE----- 443
Qy 709 KKNMADFTGAPQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPVKP 763
Db 444 --GTSNOKETKPVQAPETQINSKIANENTGEVSNKPSDKPPVEESNQPEKNGTATKP 501
Qy 764 ENS-----TDNGMLNPE---GNVGDPMDSALAEAPVDPVQEKLEKFTASYGLGDSVI 816
Db 502 ENSGNTTSNGQTEPEPSNGSTEDVSTESNTSNGNEEIKOENE-----LDPDKV 554
Qy 817 FMDGTIELR 826
Db 555 EPEKTELELR 564

RESULT 5

US-08-790-912-3
: Sequence 3, Application US/08790912
: Patent No. 5976542
: GENERAL INFORMATION:
: APPLICANT: Weiser, Jeffrey N.
: APPLICANT: Plaut, Andrew G.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
: TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: 1601 Market Street, 36th Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103-2398
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/790,912
APPLICATION NUMBER: 7600-401
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 4.4%, Score 194.5; DB 2: Length 1964;
Best Local Similarity 22.0%, Pred. No. 9.5e-06;
Matches 165; Conservative 85; Mismatches 220; Indels 281; Gaps 44;

Qy 186 GGHYHYIPKSDLSASELAALAHAGNMQPSQLSY-----SSTP 225
Db 47 GYHYKYVADSELSESE-----KKQLYDIPTYVENDDETYLVYKLNQSQ 92
Qy 226 SPSLPINPOTSKEKHE-EDGYGFDANRIIAEDSGFVMSHGDHNYFFKDLTBSQIAA 284
Db 93 LAELP-NTGSKNERCALVAGASLAALGILI-----FAVS-----KKVKNKTV--- 134
Qy 285 QKHLEEVKTSNG-----LDSLSSH-----EQDPSNAKEMKDLKKIE-EKIAGIMKOYG 334
Db 135 -LHLVLVAGMGVGLVSVHALENHLLNLYNDYELTSGEKLPLPKETSGYTYIGIREGK 193
Qy 335 VKRESIVNKKNAI-----IYPH-GDH-----HHADPIDEHKPVGIG----- 371
Db 194 TTSDPEVSNQKSAATPTKQKVOYINVTNPNFVDHSTVQAICEQTPVSTKPTVQVVEK 253
Qy 372 -----HSHSNVELEKPEGVAKREGNKVYT-----GELTNV 404
Db 254 PFSTELINPRKEEKSSDSQEQLAEHKNLET-KKEEKISPREKTGVNTLPQDEVLSQ 312
Qy 405 N-----LLKNSTFN-----NQNTLANGOKRVYSFSPPELEKLGIMN-LVKLIT-- 448
Db 313 NKPELLYREETIETKIDFOEEIQENPDIAEGTVRV-----KQEGKLGKKVEIVRFSVN 366
Qy 449 -----PDGKVLKVSCK--VFGE-----GV-----GNANFELDQPYLP 480
Db 367 KEEVSRIVSTSTTAPSPRIVEKGTGKTKQVKEQPETGVHEKDVQSGAIVEPAI-OPELP 425
Qy 481 GQTFKYTIASKDYPEVSYDGTFTVTSIAYKMASQTIFFPHAGOTYLRVNPQ-----FAV 536
Db 426 -----EAVVSDKGEPEVQP-----TLPEAVV-----TDKGET--EVQPESDTVVS 464
Qy 537 PKGTDALVRVDFEHGHNAYLENNKYVGEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAY 593
Db 465 DKGEPEQVAPLPEYKGN-----IEQVKPETPVEKTEQGEPTK--EEVPV----- 507
Qy 594 LDNQSTIYIVVEVILEKENOTDPSILPQKRN---KAQENSKLDERKVEPKTSEKV----- 646
Db 508 -----KPTETPVNPNNEGTEGTSI--QEAENPVQAEESTINSEKVSPTSSSENTGEVS 560
Qy 647 -----EKEKLSSETGNTSNTSLVEEPTVDPVQEVKVAFAESYGNKLENLVFN 693

```
Db 561 SNPSDSTSVGESNKPENHNSKENSEKTVVEVP-VNP----- 597
Qy 694 MDGTIELYLPSEVGIKKNMADFTGEAPGNGENKPSNGKV---STGTVENQPTENKRP-- 748
Db 598 NEGIVE-----GTSNQETKPKVQPAEETQTSKIANENTGEVSNKPSDSKPPV 646
Qy 749 ADSUPEAPNEKRPVKPENS-----TDNGMLNPE 775
Db 647 EESNOPEANGTATKPSNGTTSNGQTEPE 677

RESULT 6
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; City: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-4U1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-2

Query Match 4.4%; Score 194.5; DB 2; Length 2052;
Best Local Similarity 22.0%; Pred. No. 1e-05;
Matches 165; Conservative 85; Mismatches 220; Indels 281; Gaps 44;

Qy 186 GGHYHYIPKSLASASLAAAKAHLAGKNWQPSQLSY-----SSSTP 225
Db 120 GVHYKYVADSELSSSE-----KKQLVDIPTVYVNDDETYIYVYKLNQSQ 165
Qy 226 SPSLPINPTGSHEKHE-EDGYGFDPANRIIAEDSGFVNSHGCHNHYFFKDLTTEQIKAA 284
Db 166 LAELP-NTGSKSNERQALVAGASLAALGILI-----FAVS-----KKVKNKTV--- 207
Qy 285 OKHLEEVKTSNG---LDSLSH-----EODYPSNAKEMKDLKIE-EKIAIGMKQYG 334
Db 208 -LHLVLVAGMNGVLVSVALEHLLNNTDYELTSGEKLIPKEISGYIYIGYIEKG 266
Qy 335 VKRESIVVNKEKNAT-----IYPH-GDH-----HHADPIDERKHPVGIG----- 371
```

```
Db 267 TTSDFEVSNOEKSAATTTKOOKKYVDYINVTNPFVDHPSTVQAIQEQTPVSSSTKPTVEVQVKEK 326
Qy 372 -----HSHSNYELFPKEGCVAKKEGNKYVT-----GEELTNVY 404
Db 327 PFSTELINPRKEEKQSDSOEQLAEHKNLET-KKEEKISPEKXTGVTNLNPODEVLSGQL 385
Qy 405 N-----LLKNSTFN-----NQNTLANGOKRVSEFPPELEKKLGINM-LVKKLIT-- 448
Db 386 NKPELLYREETIETIKIDFQEEIQENPOLAEGTVRV-----KQEGKLGKVKVEIVRFSVN 439
Qy 449 -----PDGKVLKESVSKG--VFGB-----GV-----GNIANPELDOPILP 480
Db 440 KEEVSREIVSTSTAPRIVEKGTKKTVIKQEPETGVEHKDVQSGAIVERAI-QPELP 498
Qy 481 GQTFKYTIASKDYPEVSYDGTFTVPTSLAYKKMASQIFYPFHAGDYLRVNPQ-----FAV 536
Db 499 -----EAVVSDKGPEVQP-----TLPEAVV-----TDKGET--EYQPESPDTVVS 537
Qy 537 PKGTDALVRVDFDEHGNAYLNNYKVGEEK--LPIPKL-NQGTTRTAGNKIPVTFEMANAY 593
Db 538 DKGEPEQVAPLPEYKGN-----IEQVAPETPVKTRKQGPKEK--EEVPV----- 580
Qy 594 LDNQSTYIVVEPILEKENQTDKPSILPQFKRN---RAQENSKILDERKVEEPKTSKV--- 646
Db 581 -----KPTETPVNREGTTEGTSI--QEAENPVQPAEESTTNSKVSPTDTSSENTGEVS 633
Qy 647 -----EKEKLSGTGNTSNTSILEEYVTVDPVQEKVAKFAESYGMKLENVLN 693
Db 634 SNPSDSTSVGESNKPENHNSKENSEKTVVEVP-VNP----- 670
Qy 694 MDGTIELYLPSEVGIKKNMADFTGEAPGNGENKPSNGKV---STGTVENQPTENKRP-- 748
Db 671 NEGIVE-----GTSNQETKPKVQPAEETQTSKIANENTGEVSNKPSDSKPPV 719
Qy 749 ADSLPEAPNEKRPVKPENS-----TDNGMLNPE 775
Db 720 EESNQPEKNCTATKPSNGTTSNGQTEPE 750

RESULT 7
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.1%; Score 181.5; DB 4; Length 10182;
Best Local Similarity 19.0%; Pred. No. 0.0011;
Matches 177; Conservative 147; Mismatches 391; Indels 213; Gaps 42;

Qy 6 NQHRSOENKDNRRV---SYVDGSSQSRSKSENTPDQVSO-----KGIQAEQIVIKI 54
Db 5345 NEDSTOKTVDNAIDNGSTYITQCHNPENLKNSTIDQTSIRINTAKNDLHGVEKLQ----- 5399
Qy 55 TDGYVTSHGDHYYHNGKVPYDALFSEELMKDPNQYKLDADIVNEVG-GYIIIVKDGK 113
Db 5400 RDRGTANQELGQLGYLND--PQKS--GEESLVNGSNTRSEVEEHLNKAESLNNAMK----- 5451
```

QY	114	YYVYLKD-AAHADNVRTKDEINRQKEHVKD-NEKVSNVNAVARSQGRYTNTDGYVFENPA	171
Db	5452	--QLRDKVAEKTNYKSSDYNDSTEHQRGDDQALOEAEINIIEIGNPTLNKSEIQKL	5508
QY	172	DIITEDGNAYIPVHGHHYIPKPSDL-----SASELAARAAHLAGKNQPSOLSYS----	222
Db	5509	QQLTDAQNAL---QCSHLEEAKNNAITINKLITALNDAQBQAENVQAOQTIPAVNOQ	5565
QY	223	-----STPSPSLPINPGTSEKHEDGY-----GFDAIRIIAEDSGFVM	262
Db	5566	LTLDRINTAMQALRKVGQQNVHQQSYNFENEQPKNHYDNSVQAGQTIIDLOLDPIM	5625
QY	263	SHGDHNHYFEKDLTEODIKAAQK-HLEEVKTSH--NCGLDSLS----SHEQDYPSNAKEM	315
Db	5626	NKNEIQAQINQWTTOTALSGENKLHTDQESTNRQITGLSLTAQINAKEKDLVNOAKTR	5685
QY	316	KOLDKKI-----BEKIA-----GIMQYGVKRESIVYNKEKNAILIYPHGDDHHADPIDEH	365
Db	5686	TDMAQKLAARAKEINSAMSNLRDIGONKEDIKRSSAYINADPTKYATVDQALQNAENIINA	5745
QY	366	KPVGICHSHSNVELEKPEEGVAKBGKNVYTGELTNVVNLKNS-----TENNONFTLL	419
Db	5746	TP-----NVELNK--ATTIEQALSVOAQODLGVOOLANAKOQATQTVNGULN-SL	5793
QY	420	ANGOKRVSFSPPELEKKGILGINMLVKLTIPDGKYLEKVS-----KVFGEVGVNIAN	471
Db	5794	NDGOKR-----ELNLLINSANTRTKVOEELNKATELHNHAMEALRNSVQNVDQ	5840
QY	472	FELDQYLPQGTFKYTIASKOYPEV-SYDGTFTVPTSLAYKMASQTIFFPHAGDIYLRV	530
Db	5841	VKOSSNYV-----NEQDPQHNDN-----AVNEAQATI-----NNNAOPV	5876
QY	531	NPOFAPYKGTDALVRVDFEFGHNAYLENNYKVGSIKL-----PIPKLNQGGTT	577
Db	5877	LDKLAIERLTQVTNTKDALHGAQKLTDOQAAETGIRGLTSLNEPQKNAEYAKVTAATT	5936
QY	578	RTAGNKI---PVTFMANAYL-----DNOSTYIVE-----VPILE	608
Db	5937	RDEVNRIRQEATLTDTAMLGRKSKTKDKNFKNSKYINDHDHQOQAYDNVANNAAQOVID	5996
QY	609	KENOTDKPSILPQFRRKAQENSKL--DEKVEEPKTSK---VEKEKLSSETGSTNSNSTL	663
Db	5997	EIQATLSSDTINOLANAVTQAKSNLHGDTKLQHKDSAKQTIQALQNLNSAQKHMEDSLI	6056
QY	664	EEVPTVDPVQEKVAFESYGKHLNWLFNMDOGFIELY---LPSGEVI-----KKNMADF	715
Db	6057	DNESTRTOYQHDLTE-----AQALDGLMGALKESIKEYTIVNSNGNYINAEPSKKQAYDA	6111
QY	716	TGEAPOG--NGENKPKPEN-GKVYSTG--TVENQOPTENKPADSLPEAPH--EKVPKPNSTD	768
Db	6112	AVQNAQNIINGTNOPTINKGNVTATTQIVKTKDALDGDHLEEAKNNAQOTIRNLNLN	6171
QY	769	NGMLNPEGNVGSDPM LDSALEEAPADVQEKLE	802
Db	6172	NAQKDAEKNL-----VNSASTLEQOVQNLQ	6196

RESULT 8

```

RESOLUTION 6
US-08-242-932-2
Sequence 2, Application US/08242932
Patent No. 5595740
GENERAL INFORMATION:
APPLICANT: Brady, L. Jeannine
TITLE OF INVENTION: Cloning of No. 5595740-
the Group B Streptococcus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

```

QY 753 PEAPNEKPVKPNSTONGMLNPGNVGSDPMLDSALEEA-----PAVDPVQEKLEK 803
Db 545 VEOQDPAPI-PENSE-----MDQAKERAKIAVSKYMSKVLGDGVGHQHLQK 587
QY 804 FTASYGLGL-----DSVIFNMD 820
Db 588 KNSKIVDLFKLEAIKQOTIFDID 612
RESULT 9
US-08-714-481-2
; Sequence 2, Application US/08714481
; Patent No. 576606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 576606-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI42
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-714-481-2
Query Match 4.1%; Score 180.5; DB 1; Length 984;
Best Local Similarity 18.1%; Pred. No. 3.9e-05;
Matches 135; Conservative 109; Mismatches 246; Indels 255; Gaps 32;
QY 194 KSDLSASELAARAAHLA-----GKNQPSQLSYSTPSPSLPI-----NPG 234
Db 5 KDSVKITTEVAAPKYPYPSMAQTDGNNSSSELETTKNEIPTDDIKKAVEPVEKTAGETSA 64
QY 235 TSHEKHEEDGYGDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTS 294
Db 65 TDGKREKQLQOQKNN--LKNDYDNTILSHEQKNE--FKTKIDETNDSALLELEQFNE 120
QY 295 HNGLDLSLSHEQ--DYSNAKEMKDL--DKKIEKTAGIMKQGVKRESI-----VV 342
Db 121 TNRLLHIKQHEEVEKDKKAKQKQTLKQSDTKVD--LSNIDKELNKHQKQSVKMAEQKGIT 178
QY 343 NKEKNALIIYPHG--HHADPIDHKPVGI-----GHSNHYELFKPEEGVAKKGNKVT 396
Db 179 NEKDSMLKIEIDIRQAQQAQKDEAEVQVLEEAHSL-----KQWEDFRKFKFT 232
QY 397 GEELTNVNLKSTNNQNTFLANGQKRVSFSPPELEKKGKLGINMLVLITPDGKVLK 456

Db 233 SEQVTPKRVKRDLAANE-----NQOKIELTVSPE----- 263
QY 457 VSGKVEGEGVGNIAFELDQPYLPQOTEKYITIASDYDEVSVDGTFVTPTSLAYKMASQT 516
Db 264 -----NITVYE-----GEDVKFTVAKSDSKTILD----- 288
QY 517 IFYPHAGDTYLRVNPQFAVPAGTDAIVRVDFEFGHGNAYLNNYKVGBIKLPKLNQGT 576
Db 289 -----FSDLLTKYNPSVS-----DRISTNYKNTN---DNHKAIBITIKNLKLNESQ 331
QY 577 TRT-----AGNKIPVTIFMANAYLDNQSYTIVEVPILEKENOTDKPSILPOFKRNKAGEN 630
Db 332 TVTLKAKDDSGNVEKTF-----TIIVQKKEK-----QVPKTIPEQKD 369
QY 631 SKLDEKV-EEPKTSKVEKEKLSGTGN-----STNSNLTLEVPV----- 669
Db 370 SKTEEKVPQEPKSNKQLOELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQKSIWE 429
QY 670 ---DVOQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGVKKNHAD 714
Db 430 SOKEPIQIAITSFKKIIGDSSSKYITEHYFNKYKSDFNQYQLHAQM-----EMLTRKVVQ 484
QY 715 FTGEAPQNGENKPSKNGKST-----GTVENQPTPE-----NK-----PADSL 752
Db 485 YMNKYPDRAETIKKIFESDMKRTKEDNYSLENDALKGYFEKYFLTPTFNKIKQIVDDLDK 544
QY 753 PEAPNEKPVKPNSTONGMLNPGNVGSDPMLDSALEEA-----PAVDPVQEKLEK 803
Db 545 VEOQDPAPI-PENSE-----MDQAKERAKIAVSKYMSKVLGDGVGHQHLQK 587
QY 804 FTASYGLGL-----DSVIFNMD 820
Db 588 KNSKIVDLFKLEAIKQOTIFDID 612
RESULT 10
PCT-US95-06111-2
; Sequence 2, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville Florida
; APPLICANT: State/province: US
; APPLICANT: Country: 32611
; APPLICANT: Postal code/Zip: 904-392-8929
; APPLICANT: Phone number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.

us-09-471-255-16.ra1

Tue May 13 14:29:42 2003

```

; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI42
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-06111-2

Query Match      4.1%; Score 180.5; DB 5; Length 984;
Best Local Similarity 18.1%; Pred. No. 3.9e-05;
Matches 135; Conservative 109; Mismatches 246; Indels 255; Gaps 32;

QY 194 KDLASASALAAKHAHLA-----GKNQPSQLSYSTPSPSLPI-----NPG 234
DB 5 KDSVKITEVAAPYPSMAQTDOGNSSSELETTTMEIPTTDIKKAVEPVEKTAGTSA 64
QY 235 TSHEKHEEDYGFDARRIAEDSGVMSGHGHNHYFFKKDLTEQIKAAQKHLEEVKTS 294
DB 65 TDTCKREKQLQOWKN--LKNDVDNTILSHEOKNE--FKTKIDETNDSALLELENOFNE 120
QY 295 HNGLDLSLSHEQ-DYPSNAKEMKDL---DKKLEEKIAGIMQGVYKRESI-----VV 342
DB 121 TNRLLHIKHEEVEKDKKAAQKQTLKOSDTKYD--LSNIDKELNHOKSOVKMAEOKGIT 178
QY 343 NKEKNALIIYPHGD--HHADPIDHKKPVGI-----GHSNHYELFKPEGVAKKEGKVVY 396
DB 179 NEDKOSMLKIEDIRKQAOQADKEDAEVKVQLEEAHSLK-----KQVDEFRKKFT 232
QY 397 GEBLTNVNLLKSTFNQNTLANQKRVSFSPFPELEKGLINMLVLIPTDQGVLEK 456
DB 233 SEQVTPKRVKRLAANE-----NNQOKIELTVSPE-----263
QY 457 VSGRVFGEVGNIAFELDQPLYPGQTFKTYIAASKDYPVSYDGTFTVPTSLAYKMASQT 516
DB 264 -----NITVYE-----GEDVRETVAKSDSKTILD-----288
QY 517 IFYFPHAGDTYLRVNPQAFVKGTDALVRVDFEFGNAYLENNYKVEIKLPIPKLNQGT 576
DB 289 -----FSDLLTKYNPSVS-----DRISTNYKNT--DNHKAETITIKNLKLSQ 331
QY 577 TRT-----AGNKIPVTFMANAYLDNOSTVIVBVPFILEKENOTDKPSILPQPKRNKAQEN 630
DB 332 TVTLKAKDDSGNVVEKTF-----TITVQKKEK-----QVPKTPQKXD 369
QY 631 SKLDEKV-EEPKTSEKVEKELSETGN-----STNSILEEVPV-----669
DB 370 SKTEKVPQEPKSDKNQLOELIKSAQOOLEKLEKATKELMEQPEIPSNPEYGIQKSIWE 429
QY 670 --DPVQEKVAKP-----AESYCKLENVLFNMDGTIELYLPSPGEVIKKNMAD 714
DB 430 SQKEPQIEATISFKKIIGDSSSKYTYTHYFNKYKSPFMVQLHAQM-----EMLTKEVQV 484
QY 715 FTGAPOGNGENKPSNGKYST-----GIVENQTE-----NK-----PADSL 752
DB 485 YMKYVPDAEIKKIFESDMKRTKEDNYGSLNDALKGYPEKYEYLTFFNKIKQIVDLDKK 544
QY 753 PEAPNEKVPKPNSTDNGLNPNEGVSDPWLDSALEEA-----PAVDPVQEKLEK 803
DB 545 VEQDPAPPI-PENSE-----MDQAREKAKIAVSKYMSKVLQGVGHOHLQK 587
QY 804 FTASYGLGL-----DSVIFNMD 820
DB 588 KNSKIVDLKLELAIKQOIFDID 612

RESULT 11
US-08-296-791-6
; Sequence 6, Application us/08296791

```

```

; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

Query Match      4.1%; Score 180.5; DB 4; Length 1848;
Best Local Similarity 17.9%; Pred. No. 9.9e-05;
Matches 165; Conservative 128; Mismatches 345; Indels 283; Gaps 35;

QY 12 ENKNNRYSYVDGSG--SSQKSENLTPOVSOKEGIAQBIQIKITDQGYVISHGDVHYHY 70
DB 470 GKNEGLKLVGGDTVILKOKADANNKQAFSQVGLVSGRSTLVLNDDKQVDPNPIYFGR 529
QY 71 NGKVPYDA-----LFSEELMKDPN-----YOLKDAI 98
DB 530 GORLDLNGSLTFDHIRNTDDGARVVNHNMTNTSNTITGESLIINPTITITSYNIEAQDD 589
QY 99 VNEV-----KGGHIIKVDGKYVYLKDAHADNV-----RTKDEINRQ 136
DB 590 DHPLEIRSIPIYQLFENQDNKRSYTLKKGASTRSELPQNSGESNENWLYMGRTSDAKRN 649
QY 137 KCEHVKNDEKVASNVAVARSQGRVTTNDGYFENPADIIEDTGNNAVIVPHGGHYHYIPKSD 196
DB 650 VVNHII--NERNNG-----FNGYGEETKATQNGKLVNTENG-----KSD 688
QY 197 LSASELAARAKAHLAGKNQPSQLSYSTPSPSLPNTGTSHEKHEEDYGFDANRIIAD 256
DB 689 QNRFLITGNTLNGDLNVEKGTFLSGRPTPHARDIAGISSTK--KDPHFTENNEVVVED 746
QY 257 E-----SGFVMSHGDHNYFFKKDLTEQIKAAQKH 287
DB 747 DMINRFKATTVNTGNASLYSGRVNANITSNTASNNAQVHIGYK---TGDTCVRSYD 803
QY 288 LEEVKTSHNGLDLSLSHEQDYPNAKEMKDLKKI-----BEKTAGIMKQGVKRESIVV 342
DB 804 TGVV-TCHNSNLSEKALNSFNPTNLRGVNVNTENASFTLGRANLFGTISQIGTSQ----V 858
QY 343 NKEKNALIIYPHGDHHAADPIDHKKPVGI--GHSNHYELFKPEGVAKKEGKVVYTGSEL 400
DB 859 NLKENS-----HWHLTGNSVNVNQLNLTNGHILN-----AONDANKVITYNTL 901

```

```

QY 401 TNNVLLKNSTF-----NNQNFLLANGQKRVSFSPPELEKKGINMLVLTLPDGK 452
Db 902 T-VNSLSGNGSYVWVDFTNKSNKVVN--KSATGNTLQVADKTGFENHNELTLFDAS 958
QY 453 VLEKVGKRVGCGVGNIAFELDQPLFGQTFKYTIASKD-----YPEVSYDGTFTVPT 506
Db 959 NATRNLEV-----TLANGSVDR-----GAWKKYKLRNVNGRYDLYNPEV-----997
QY 507 SLAYKMASOTIFYPFHAGDTYLRVNP---QFAPVPGTDLVRVDFEFHGNAYLENNYKVG 563
Db 998 ---EKRNTV-----DTNITTPNDIQADAPSAQ-----SNNEEIA 1030
QY 564 EIKLPIKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVVEPILEKENOTDKPSILPOFK 623
Db 1031 RVETVPPPPAPATSAIASEQPET-----RPAETAQPA-----1063
QY 624 RNKAQENSKLDEKVEEPTSEKVEKELSETGNTSNTLEEVPVDPVQEKVAKFAESY 683
Db 1064 -----MEETNTANSTETAPKSDTATQTENPENSEVPS-----ETTEKVAENP 1105
QY 684 GMLKLVLFNMDGTIELYLPSEGVTKKNMADF-----TGEAPQNGENKPSGNGKVSTGT 738
Db 1106 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQ-----TAE 1160
QY 739 VENOPT-----ENKPADSLPEAPNEKPV-----KPNSTDNGLNPEGNVGSOPM-----783
Db 1161 TKSEPTESVTVSENOPEKTVSQSTEDKVVVEKEERAKVETEETQKAPQVTSKEPPKQAE 1220
QY 784 -----LDSALEEAPAVDPVQ 798
Db 1221 APEEVPDTNABEAQALQQTQ 1241

```

RESULT 12

```

PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RET
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid

```

TOPOLOGY: unknown

PCT-US95-10661A-6

Query Match 4.1%; Score 180.5; DB 5; Length 1848;

Best Local Similarity 17.9%; Pred. No. 9.9e-05;

Matches 165; Conservative 128; Mismatches 345; Indels 283; Gaps 35;

```

QY 12 ENKONNRVSYVDGSQ--SSQKSENLTDPQVQKEGIAEQIVIKITDQGYVTSHGHDHYHY 70
Db 470 KGKNEGLLVKVDGIVLKQKADANNKQVAFSGVIGSRSTLVLDNDQVDPNSIFGFR 529
QY 71 NGKVPYDA-----LFSEELMKDPN-----YOLKDAADI 98
Db 530 GGRDLNGNSLTFDIRNIDDGARVVNHNMTNITSNITITGESLTNPNTITSYNIEAQDD 589
QY 99 VNEY-----KGGYIILKVDGKYVYLKDAHADNV-----RTKDEINRQ 136
Db 590 DHPRIIRISYRQLYFNQDNRSYITLKGASTSELSQPNSGESNENWLYWGRTSDAAKRN 649
QY 137 KOEHVKONEKYNVAVARSQGRYTTNDGVYFNPADIIDTGNAYIVPHGHGHYHYIPKSD 196
Db 650 VMNHI--NNERNMG-----FNGYFGEETKATQNGKLNVTFG-----KSD 688
QY 197 LSASELAAAKAHLAGKNMOPSQLSYSTPSPSLPINPTGSHEKHEEDGYGFDANRIIAED 256
Db 689 QNRPELLTGGTNGLDNLVEKGTIFLSCRPTPHARDIAGISSTK--KDPHTTENNEVVVED 746
QY 257 E-----SGFVMSHGDHNYHFFKKDLTDEQIKAAQKH 287
Db 747 DWINRNEKATTMNVTGNASLYSGRNVANITSNITASNAQVHGYK---TGDVTCVRSDY 803
QY 288 LEEVKTSHNGLDLSLSEHQDYPSNAKEMKOLDKKI-----EETIAGIMKQYGVKRESIV 342
Db 804 TGYV--TCHNSMLSEKALNSFNPTNLRGNVNLNENASTLTKANLFGTIQIGTSQ-----V 858
QY 343 NKEKNAIYPHGDHHDADPIDEHKPVGI--CHSHSNTELEKPEBEGVAKKCGNKVYTGEL 400
Db 859 NLKENS-----HWHLTGSNVNQLNLTNGHILN-----AQNDANKVTTVNTL 901
QY 401 TNNVLLKNSTF-----NNQNFLLANGQKRVSFSPPELEKKGINMLVLTLPDGK 452
Db 902 T-VNSLSGNGSYVWVDFTNKSNKVVN--KSATGNTLQVADKTGFENHNELTLFDAS 958
QY 453 VLEKVGKRVGCGVGNIAFELDQPLFGQTFKYTIASKD-----YPEVSYDGTFTVPT 506
Db 959 NATRNLEV-----TLANGSVDR-----GAWKKYKLRNVNGRYDLYNPEV-----997
QY 507 SLAYKMASOTIFYPFHAGDTYLRVNP---QFAPVPGTDLVRVDFEFHGNAYLENNYKVG 563
Db 998 ---EKRNTV-----DTNITTPNDIQADAPSAQ-----SNNEEIA 1030
QY 564 EIKLPIKLNQGTTRTAGNKIPVTMANAYLDNQSTYIVVEPILEKENOTDKPSILPOFK 623
Db 1031 RVETVPPPPAPATSAIASEQPET-----RPAETAQPA-----1063
QY 624 RNKAQENSKLDEKVEEPTSEKVEKELSETGNTSNTLEEVPVDPVQEKVAKFAESY 683
Db 1064 -----MEETNTANSTETAPKSDTATQTENPENSEVPS-----ETTEKVAENP 1105
QY 684 GMLKLVLFNMDGTIELYLPSEGVTKKNMADF-----TGEAPQNGENKPSGNGKVSTGT 738
Db 1106 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQ-----TAE 1160
QY 739 VENOPT-----ENKPADSLPEAPNEKPV-----KPNSTDNGLNPEGNVGSOPM-----783
Db 1161 TKSEPTESVTVSENOPEKTVSQSTEDKVVVEKEERAKVETEETQKAPQVTSKEPPKQAE 1220
QY 784 -----LDSALEEAPAVDPVQ 798
Db 1221 APEEVPDTNABEAQALQQTQ 1241

```

RESULT 13

TITLE OF INVENTION: BINDING DOMAINS FROM LERKININ AND THEIR USE IN THE TREATMENT OF BLOOD DISEASES


```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-4

Query Match 4.1%; Score 180; DB 2; Length 1435;
Best Local Similarity 19.3%; Pred. No. 7.4e-05;
Matches 113; Conservative 95; Mismatches 197; Indels 182; Gaps 26;

QY 11 QENKDNRRVSYVDGSSQSSQSENLTDP-----QVSQKEGIAEQIIV---KITDQGYVTSH 63
Db 789 QKERDDSLSKISVSPNSRPETDAKDTSNLLKLGQVDISMPKAVIGSSPNDINVTQ 848
QY 64 GDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEY----- 102
Db 849 GDNISGVNKPDLSDVDPDKKELEDQNSDESEETVVNHISKSPINNGDDSGSGSATVSE 908
QY 103 -----KGGYIIKVD--GKYVYVYLKDAHADNVRTKDEINRQKQEHVKDNEKVNNAVARS 156
Db 909 SSSSNTGLSDDDRRNGDTFVRTQDTANTEDVIRKENADKDEKGADEHERHSTSESLSPP 968
QY 157 QGRVTTND--GYVFNPAIIEDTGN--YIVPHGGHYHYIPKSDLS-----ASELAATA 207
Db 969 EEKMLTDNEGNSLNHEEYKHEHTSNSDNVQSGGIYVNMVNEKELKDTLENPSSSLDEGKA 1028
QY 208 H--LAGKNMQPSQLSYSTSPSPSLPINPGTSHE----- 238
Db 1029 HEELSEPNLSSDQ--DMSNTPGPL-----DNTSEETTERISNNEYKVNREDERTLKEYED 1083
QY 239 -----KHEEDGYGFDANRIIA--EDESGFV-----MSHGDHNYHFF---KK 274
Db 1084 IYLVKSHMNRSDGELYDENSLSVTNDESDAEAKMKGNDTSEMNSHNSQHIESDQOKN 1143
QY 275 DL-----TEEQIAAAQK-----HLEEVKTSHNGLDSSSHEQDY 308
Db 1144 DMKTVGDLGTHVQNEISVPVTGEIDEKLRKESKSIHKAEEELSHTDIHKI--NEEDR 1201
QY 309 PSNAKEMKDLKTEEKIAGIMKQYGVKRISIVYVNEKKNALIIYPHGDH-----HADPIDE 364
Db 1202 NSNTLHLKDIRNEENER-----HLTNQINISQERD--LQKHGFHTMNLHGDGVSE 1251
QY 365 HRPVGIHSHSNYELFKPEGVAKKEGKVYTGEBELTNVNLKNTSTNNQNTFLANGQK 424
Db 1252 RSQ--INSH-----HGNRQDRGSGNVLNMR-----NNNNFN--NIPS 1288
QY 425 RVSFSPPELEKKLGINM-----LVKLITPDGKVLKVSQK 460
Db 1289 RYNL-----YDKKLDLDLYENRNDSTTKELIKLAEINKCENEISVK 1330

RESULT 15
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESS: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
```

```
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPL1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-4

Query Match 4.1%; Score 180; DB 2; Length 1435;
Best Local Similarity 19.3%; Pred. No. 7.4e-05;
Matches 113; Conservative 95; Mismatches 197; Indels 182; Gaps 26;

QY 11 QENKDNRRVSYVDGSSQSSQSENLTDP-----QVSQKEGIAEQIIV---KITDQGYVTSH 63
Db 789 QKERDDSLSKISVSPNSRPETDAKDTSNLLKLGQVDISMPKAVIGSSPNDINVTQ 848
QY 64 GDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEY----- 102
Db 849 GDNISGVNKPDLSDVDPDKKELEDQNSDESEETVVNHISKSPINNGDDSGSGSATVSE 908
QY 103 -----KGGYIIKVD--GKYVYVYLKDAHADNVRTKDEINRQKQEHVKDNEKVNNAVARS 156
Db 909 SSSSNTGLSDDDRRNGDTFVRTQDTANTEDVIRKENADKDEKGADEHERHSTSESLSPP 968
QY 157 QGRVTTND--GYVFNPAIIEDTGN--YIVPHGGHYHYIPKSDLS-----ASELAATA 207
Db 969 EEKMLTDNEGNSLNHEEYKHEHTSNSDNVQSGGIYVNMVNEKELKDTLENPSSSLDEGKA 1028
QY 208 H--LAGKNMQPSQLSYSTSPSPSLPINPGTSHE----- 238
Db 1029 HEELSEPNLSSDQ--DMSNTPGPL-----DNTSEETTERISNNEYKVNREDERTLKEYED 1083
QY 239 -----KHEEDGYGFDANRIIA--EDESGFV-----MSHGDHNYHFF---KK 274
Db 1084 IYLVKSHMNRSDGELYDENSLSVTNDESDAEAKMKGNDTSEMNSHNSQHIESDQOKN 1143
QY 275 DL-----TEEQIAAAQK-----HLEEVKTSHNGLDSSSHEQDY 308
Db 1144 DMKTVGDLGTHVQNEISVPVTGEIDEKLRKESKSIHKAEEELSHTDIHKI--NEEDR 1201
QY 309 PSNAKEMKDLKTEEKIAGIMKQYGVKRISIVYVNEKKNALIIYPHGDH-----HADPIDE 364
Db 1202 NSNTLHLKDIRNEENER-----HLTNQINISQERD--LQKHGFHTMNLHGDGVSE 1251
QY 365 HRPVGIHSHSNYELFKPEGVAKKEGKVYTGEBELTNVNLKNTSTNNQNTFLANGQK 424
Db 1252 RSQ--INSH-----HGNRQDRGSGNVLNMR-----NNNNFN--NIPS 1288
QY 425 RVSFSPPELEKKLGINM-----LVKLITPDGKVLKVSQK 460
Db 1289 RYNL-----YDKKLDLDLYENRNDSTTKELIKLAEINKCENEISVK 1330
```

Tue May 13 14:29:42 2003

us-09-471-255-16.ra1

Page 12

Search completed: May 13, 2003, 13:58:10
Job time : 31.0205 secs

Db 181 YIVPHGGHYIIPKSDLSASELAARAHLAGKNQPSQLSYSTSPSPINPGTSHEKH 240
Qy 241 BEDGYFDANRIIAEDSGFVMSHGDNHYFFKKDLTEQOIKAAQKHLSEVKTSHNGLDS 300
Db 241 BEDGYFDANRIIAEDSGFVMSHGDNHYFFKKDLTEQOIKAAQKHLSEVKTSHNGLDS 300
Qy 301 LSSHEQDYPNAKEMKDLKIEEKIAGIMQYGVRESIVVNKEKNAIIPYHGDHHDAD 360
Db 301 LSSHEQDYPNAKEMKDLKIEEKIAGIMQYGVRESIVVNKEKNAIIPYHGDHHDAD 360
Qy 361 PIDCHKVPGIGHSHSNVELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNFLLA 420
Db 361 PIDCHKVPGIGHSHSNVELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNFLLA 420
Qy 421 NGKRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLP 480
Db 421 NGKRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLP 480
Qy 481 GQPKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFPFHAGDTYLRVNPQFAVPKGT 540
Db 481 GQPKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFPFHAGDTYLRVNPQFAVPKGT 540
Qy 541 DALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTY 600
Db 541 DALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTY 600
Qy 601 IVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 660
Db 601 IVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 660
Qy 661 STLSEVPTDVPQVKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVYKKNMADTFGEAP 720
Db 661 STLSEVPTDVPQVKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVYKKNMADTFGEAP 720
Qy 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLMPEGNVGS 780
Db 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLMPEGNVGS 780
Qy 781 DPMLDLSALEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVYKKNLLIS 840
Db 781 DPMLDLSALEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVYKKNLLIS 840

RESULT 2

US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 035190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 97.38; Score 4276.5; DB 9; Length 1039;
Best Local Similarity 82.38; Pred. No. 1.4e-245;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

Qy 1 CAYALNQHRSEKNDNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYV 60
Db 21 CAYALNQHRSEKNDNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYV 80
Qy 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGGYIKVDGKYVYVLKD 120
Db 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGGYIKVDGKYVYVLKD 140
Qy 121 RAHADNVTKDEINPQKQEHVKDNKNSNVAVARSQGRYTTNDGYVFNPAADITDIGNA 180
Db 141 RAHADNVTKDEINPQKQEHVKDNKNSNVAVARSQGRYTTNDGYVFNPAADITDIGNA 200
Qy 181 YIVPHGGHYHYPKSDLSASELAARAHLAGKNQPSQLSYSTASDNNQTSQVAKGSTK 223
Db 201 YIVPHGGHYHYPKSDLSASELAARAHLAGKNQPSQLSYSTASDNNQTSQVAKGSTK 260
Qy 224 ----- 223
Db 261 PANKSENLSQLLKELYDSPAQRYSSEDLVDFPAKIIISRTPNGVAIPHGDHYHIFPYSK 320
Qy 224 ----- 223
Db 321 LSALEKTAARVPISTGTSVTNAKPNVSVSLGSLSSNPSSLTTSKELSSASDGYIFN 380
Qy 224 -----TPSPSLPINPOTSHEKHEED 243
Db 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPOTSHEKHEED 440
Qy 244 GYGFANRIIAEDSGFVMSHGDNHYFFKKDLTEQOIKAAQKHLSEVKTSHNGLDSLS 303
Db 441 GYGFANRIIAEDSGFVMSHGDNHYFFKKDLTEQOIKAAQKHLSEVKTSHNGLDSLS 500
Qy 304 HQDQXPSNAKEMKDLKIEEKIAGIMQYGVRESIVVNKEKNAIIPYHGDHHDADID 363
Db 501 HQDQXPSNAKEMKDLKIEEKIAGIMQYGVRESIVVNKEKNAIIPYHGDHHDADID 560
Qy 364 EHKPVGIGHSHSNVELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNFLLA 423
Db 561 EHKPVGIGHSHSNVELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNFLLA 620
Qy 424 KRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLPQGT 483
Db 621 KRVSFSPPELEKKGILGNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLPQGT 680
Qy 484 FKYTITASKDYPEVSYDGTFTVPTSLAYKMASQTIIFPFHAGDTYLRVNPQFAVPKGTAL 543
Db 681 FKYTITASKDYPEVSYDGTFTVPTSLAYKMASQTIIFPFHAGDTYLRVNPQFAVPKGTAL 740
Qy 544 VRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVE 603
Db 741 VRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVE 800
Qy 604 VPILKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 663
Db 801 VPILKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 860
Qy 664 EEPVTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVYKKNMADTFGEAPOGN 723
Db 861 EEPVTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVYKKNMADTFGEAPOGN 920
Qy 724 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLMPEGNVGSDFM 783
Db 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLMPEGNVGSDFM 980
Qy 784 LLSALEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVYKKNL 837
Db 981 LLSALEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVYKKNL 1034

RESULT 3

US-09-884-465A-379
; Sequence 379, Application US/09884465A
; Publication No. US20030077293A1

```

: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 379
: LENGTH: 1152
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
: NAME/KEY: MISC_FEATURE
: LOCATION: (1)..(1)
: OTHER INFORMATION: xaa = Methionine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (344)..(344)
: OTHER INFORMATION: xaa = Glycine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (345)..(345)
: OTHER INFORMATION: xaa = Proline or nothing
: US-09-884-465A-379

Query Match          77.4%; Score 3402.5; DB 9; Length 1152;
Best Local Similarity 79.6%; Pred. No. 1.e-193;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SOENKNNRVSYVDGQS--SQKSENL-----TPDOVSQKEGTQAEQI-VIKIT 55
DB 355 SSTAQNNQTSQVAKGSTPANKSENLSLLKELXDPSAQYSDGLVDFDPKAIISRT 414
QY 56 DQGYVTSBGHDHYHYNGKVPYDALFS-EELLMKDPNYQLKDADIVNEVKGVIIVDGY 114
DB 415 PNGVAIPHGDHYHF-----IPYKLSALEKTIAR-----MVPISGT- 450
QY 115 YVYLKDAADNVRTKDEINROKQEHVKNEKNSVAVARSQGRYTT-----NDGYV 167
DB 451 -----GSTVST-----NAKPNYVSSLGSLSSNPSSLTTSKELSSASDGYI 491
QY 168 FNPADIETGNAYIVPHGCHYHYIPKSDLSASELAAGKAGNMOPSQLSYS-STPS 226
DB 492 FNPKDIVEETATAYIVRHGDHPHYIPKSNQIG-----QTLFNNSLATPS 536
QY 227 PSLPINPGTSHKEHEDGCGFDANRTIAEDSGFVNSHGDHNYHFKKDLTEEQIKAAQK 286
DB 537 PSLPINPGTSHKEHEDGCGFDANRTIAEDSGFVNSHGDHNYHFKKDLTEEQIKAAQK 596
QY 287 HLEEVKTSUNGSLDSSSHEDQYPSNAKEMKDKLTKTEEKIAGIMKOYGVKRESIVVNEK 346
DB 597 HLEEVKTSUNGSLDSSSHEDQYPSNAKEMKDKLTKTEEKIAGIMKOYGVKRESIVVNEK 656
QY 347 NAIIVPHGDHHDPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVKYVTGELTNVYNL 406
DB 657 NAIIVPHGDHHDPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKVKYVTGELTNVYNL 716
QY 407 LKNSIFNNQNTLANGQKRVSFSPPELEKLGINMLVKLIITPDGKVLKVKSGKVGFGV 466
DB 717 LKNSIFNNQNTLANGQKRVSFSPPELEKLGINMLVKLIITPDGKVLKVKSGKVGFGV 776
QY 467 GNIAFELDQYLPFGTFFKTYTIAKSDYPEVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 526
DB 777 GNIAFELDQYLPFGTFFKTYTIAKSDYPEVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 836
QY 527 YLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPV 586
DB 837 YLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPV 896
QY 587 TFMANAYLDNQSTYIVVEPILKEKNTQKPSLTPQFKRKAQENSKLDEKVEEPTSEKV 546
DB 897 TFMANAYLDNQSTYIVVEPILKEKNTQKPSLTPQFKRKAQENSKLDEKVEEPTSEKV 956
QY 647 EKEKLSSETGNSNSTLEEVPTVDPVQEKVAFABSYGKMLNVLNFMMDGTIELYLPSE 706
DB 957 EKEKLSSETGNSNSTLEEVPTVDPVQEKVAFABSYGKMLNVLNFMMDGTIELYLPSE 1016
QY 707 YTKKNWADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPADSLPEAPNEKPKVPKPS 766
DB 1017 YTKKNWADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPADSLPEAPNEKPKVPKPS 1076
QY 767 TDNGMLNPEGNVSGDPMELDSALEEAPVDPVQEKLEKFTASYGLDLSVIFNMDGTIELR 826
DB 1077 TDNGMLNPEGNVSGDPMELDSALEEAPVDPVQEKLEKFTASYGLDLSVIFNMDGTIELR 1136
QY 827 LPSGEVIRKKNL 837
DB 1137 LPSGEVIRKKNL 1147

RESULT 4
US-09-884-465A-381
: Sequence 381, Application US/09884465A
: Publication No. US20030077293A1
: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 381
: LENGTH: 1238
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
: NAME/KEY: MISC_FEATURE
: LOCATION: (1)..(1)
: OTHER INFORMATION: xaa = Methionine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (430)..(430)
: OTHER INFORMATION: xaa = Glycine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (431)..(431)
: OTHER INFORMATION: xaa = Proline or nothing
: US-09-884-465A-381

Query Match          77.4%; Score 3402.5; DB 9; Length 1238;
Best Local Similarity 79.6%; Pred. No. 1.e-193;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SOENKNNRVSYVDGQS--SQKSENL-----TPDOVSQKEGTQAEQI-VIKIT 55
DB 441 SSTAQNNQTSQVAKGSTPANKSENLSLLKELXDPSAQYSDGLVDFDPKAIISRT 500
QY 56 DQGYVTSBGHDHYHYNGKVPYDALFS-EELLMKDPNYQLKDADIVNEVKGVIIVDGY 114
DB 501 PNGVAIPHGDHYHF-----IPYKLSALEKTIAR-----MVPISGT- 536
QY 115 YVYLKDAADNVRTKDEINROKQEHVKNEKNSVAVARSQGRYTT-----NDGYV 167
```


Dd 1290 TDNGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTTASYGLGLDSVIFNMDGTIELR 1349

Qy 827 LPSGEVIKKNL 837

Dd 1350 LPSGEVIKKNL 1360

RESULT 6

US-09-884-465A-378

; Sequence 378, Application US/09884465A

; Publication No. US20030077293A1

; GENERAL INFORMATION:

; APPLICANT: Shire Biochem, Inc.

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard

; APPLICANT: Martin, Denis

; APPLICANT: Charland, Nathalie

; APPLICANT: Ouellet, Catherine

; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044

; CURRENT APPLICATION NUMBER: US/09/884,465A

; CURRENT FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: 60/212,683

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 378

; LENGTH: 1378

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Unknown Organism

; NAME/KEY: MISC_FEATURE

; LOCATION: (1)..(1)

; OTHER INFORMATION: Xaa = Methionine or nothing

; NAME/KEY: MISC_FEATURE

; LOCATION: (570)..(570)

; OTHER INFORMATION: Xaa = Glycine or nothing

; NAME/KEY: MISC_FEATURE

; LOCATION: (571)..(571)

; OTHER INFORMATION: Xaa = Proline or nothing

; US-09-884-465A-378

Query Match

Best Local Similarity 77.4%; Score 3402.5; DB 9; Length 1378;

Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

Qy 10 SQENKDNVSVYDGSQS--SOKSENL-----TPDQVSQKEGIAQEI-VIKIT 55

Dd 581 SSTASDNTQSVAKGSTKPAKSENQLQLKELYDSPAQRYSDDGLVFPDAKIIST 640

Qy 56 DOGYVTSBGHDHYHYNGKYPYDALFS--EELLMKDPNYQLKDADIVNEVKGYYIRKVDGKY 114

Dd 641 PNGVAIPHGDHYHF-----IPYSKLSALEEKIAR-----MVPISGT- 676

Qy 115 YVYLKDAHADNVRIKDEINRQKQEVKDNKVNVAVARSQGRYTT-----NDGYV 167

Dd 677 -----GSTVST-----NAKNEVYVSSGLSSNPSLTTSKELSSASDGYI 717

Qy 168 FNPADIIETGNAYIVPHGGHYHYIPKSDLSASELAAKAHLAKGNWOPSOVSYS-STPS 226

Dd 718 FNPKDIVEETATAYIVRHGDHFHYIPKSNQIG-----QPTLPNNSLATPS 762

Qy 227 PSLPDPNGTSHKHEEDGFGFDPAKRIIAEDSGFVMSHGDHNYHFFKDLTEEQIKAAQK 286

Dd 763 PSLPDPNGTSHKHEEDGFGFDPAKRIIAEDSGFVMSHGDHNYHFFKDLTEEQIKAAQK 822

Qy 287 HLEEVKTSNGDLSLSSHEODYPNAKMKDLKXIEKIAKIMKOYGVKRESIVVNKEK 346

Dd 823 HLEEVKTSNGDLSLSSHEODYPNAKMKDLKXIEKIAKIMKOYGVKRESIVVNKEK 882

Qy 347 NAIITYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGELTNVNL 406

Dd 1290 TDNGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTTASYGLGLDSVIFNMDGTIELR 1349

Qy 827 LPSGEVIKKNL 837

Dd 1350 LPSGEVIKKNL 1360

Dd 883 NAIITYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGELTNVNL 942

Qy 407 LKNSTFNQNFHTLANGQKRVSFSPPELEKLGINMLVKLITPDGKVLKESGVKFGGCV 466

Dd 943 LKNSTFNQNFHTLANGQKRVSFSPPELEKLGINMLVKLITPDGKVLKESGVKFGGCV 1002

Qy 467 GNIAFELDQPYLPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDT 526

Dd 1003 GNIAFELDQPYLPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDT 1062

Qy 527 YLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNOGTTTAKNKTIV 586

Dd 1063 YLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNOGTTTAKNKTIV 1122

Qy 587 TFMANAYLDNQSTYIIVEYPILEKENQTDKPSILPOFKENKAQENSKLDEKVEEPEKTSKV 646

Dd 1123 TFMANAYLDNQSTYIIVEYPILEKENQTDKPSILPOFKENKAQENSKLDEKVEEPEKTSKV 1182

Qy 647 EKEKLSSETGNSNSTLEEVTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSE 706

Dd 1183 EKEKLSSETGNSNSTLEEVTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSE 1242

Qy 707 VIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGTVENQPTENKPADSLPEAPNEKPVKPENS 766

Dd 1243 VIKKNMADFTGEAPOGNGENKPSSENGKYSTGTGTVENQPTENKPADSLPEAPNEKPVKPENS 1302

Qy 767 TDNGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTTASYGLGLDSVIFNMDGTIELR 826

Dd 1303 TDNGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTTASYGLGLDSVIFNMDGTIELR 1362

Qy 827 LPSGEVIKKNL 837

Dd 1363 LPSGEVIKKNL 1373

RESULT 7

US-09-884-465A-370

; Sequence 370, Application US/09884465A

; Publication No. US20030077293A1

; GENERAL INFORMATION:

; APPLICANT: Shire Biochem, Inc.

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard

; APPLICANT: Martin, Denis

; APPLICANT: Charland, Nathalie

; APPLICANT: Ouellet, Catherine

; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044

; CURRENT APPLICATION NUMBER: US/09/884,465A

; CURRENT FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: 60/212,683

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 370

; LENGTH: 906

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Unknown Organism

; US-09-884-465A-370

Query Match

Best Local Similarity 73.6%; Score 3237; DB 9; Length 906;

Matches 658; Conservative 35; Mismatches 82; Indels 120; Gaps 12;

Qy 25 SQSQSKSENLPDQVSQKEGIAQEI-----QIVIKITDQG 58

Dd 45 ACAVAKERGLTPPTDQDSDGNTAEKGAEEIYNRVKAAKVPDLPDRMPYNLYQTVTEVAKGS 104

Qy 59 YVTSBGHDHYHYNGKYPYDALFS--EELLMKDPNYQLKDADIVNEVKGYYIRKVDGKYYVWL 118

Dd 105 LIIPHGDHYH--NKKFEW---FDEGLYEAPKGSLED-----LLATVYKIVEP 147

QY 119 KDAHADNVRTKDEINROKQERHVDNEKVNNAVARSQGRYTTNDGYVFNPAADIIEDTG 178
Db 148 RNA--SDHVRK-----NKADQSKPDEKDEHEV-----SETHPESD-----EKEN 187
QY 179 NAYIVPHGCHYHYTPKSDLSASELAAA-----KAHLAG----- 211
Db 188 HAGLPSADNL-YKPSDTDETEEEAEDTTDEAIPGTPSIRONAMETTLGLKSSLLGT 246
QY 212 -----KNMPSQL-----SYSTSPSPSLPNNSLATPSPLPNNPSTNS 242
Db 247 KDNNTISAEDVSLALLKESQAPIQGPQIGQPTLPNNSLATPSPLPNNPSTNS 306
QY 243 DGYGFANDRIIAEDSGFVNSHGDHNYFFPKDLTEQIKAAQKHLEEVKTSNGLDLS 302
Db 307 DGYGFANDRIIAEDSGFVNSHGDHNYFFPKDLTEQIKAAQKHLEEVKTSNGLDLS 366
QY 303 SHEQDYPNAKEMKDLKIKIEKTAGIMKQYGVKRESIVVYVNEKKNALIIYPHGDHHDPI 362
Db 367 SHEQDYPNAKEMKDLKIKIEKTAGIMKQYGVKRESIVVYVNEKKNALIIYPHGDHHDPI 426
QY 363 DEHKPWGICHSHSNYELFKPEEGYAKKEGKNVYTGELTNVYVLLKNSTFNQNTFLANG 422
Db 427 DEHKPWGICHSHSNYELFKPEEGYAKKEGKNVYTGELTNVYVLLKNSTFNQNTFLANG 486
QY 423 QKRVSFSPPELEKKGILNMLVKLITPDGKVLKESVKGVEGEGVGNIANFELDQYLPQ 482
Db 487 QKRVSFSPPELEKKGILNMLVKLITPDGKVLKESVKGVEGEGVGNIANFELDQYLPQ 546
QY 483 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDA 542
Db 547 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDA 606
QY 543 LVRVDFEFGHGNAYLNNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIV 666
QY 603 EYPTLEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTSEKVEKELSETGNSTNS 662
Db 667 EYPTLEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTSEKVEKELSETGNSTNS 726
QY 663 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQ 722
Db 727 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQ 786
QY 723 NGENKPSNGKVSIGTGTVENQPTENKPADSLPEAPNPKVPKPNSTNGMLNPEGVNSDP 846
QY 783 MLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVIRKKNL 837
Db 847 MLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVIRKKNL 901

RESULT 8

US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332

Query Match 73.4%; Score 3228; DB 9; Length 906;
Best Local Similarity 73.4%; Pred. No. 1.6e-183;
Matches 657; Conservative 35; Mismatches 83; Indels 120; Gaps 12;

QY 25 SOSKOKSENITPDQVSKQKEGIAE-----QIVIKITDQ 58
Db 45 AQAYAKERGLTPPTSDHDSNGTEAKGAEAIYNRVKAARKVPLDRMPYNQLOXTVEKNGS 104
QY 59 YVTSHGDRHYHYNGKVPYDALFSEELMKDPNYPQLDADIVNEVGKGYIKRVGKYVYVL 118
Db 105 LIIPSYDHYH--NIKFEW--FDEGLYEAAPKGYSLD-----LLATVKYVVEP 147
QY 119 KDAHADNVRTKDEINROKQERHVDNEKVNNAVARSQGRYTTNDGYVFNPAADIIEDTG 178
Db 148 RNA--SDHVRK-----NKADQSKPDEKDEHEV-----SETHPESD-----EKEN 187
QY 179 NAYIVPHGCHYHYTPKSDLSASELAAA-----KAHLAG----- 211
Db 188 HAGLPSADNL-YKPSDTDETEEEAEDTTDEAIPGTPSIRONAMETTLGLKSSLLGT 246
QY 212 -----KNMPSQL-----SYSTSPSPSLPNNSLATPSPLPNNPSTNS 242
Db 247 KDNNTISAEDVSLALLKESQAPIQGPQIGQPTLPNNSLATPSPLPNNPSTNS 306
QY 243 DGYGFANDRIIAEDSGFVNSHGDHNYFFPKDLTEQIKAAQKHLEEVKTSNGLDLS 302
Db 307 DGYGFANDRIIAEDSGFVNSHGDHNYFFPKDLTEQIKAAQKHLEEVKTSNGLDLS 366
QY 303 SHEQDYPNAKEMKDLKIKIEKTAGIMKQYGVKRESIVVYVNEKKNALIIYPHGDHHDPI 362
Db 367 SHEQDYPNAKEMKDLKIKIEKTAGIMKQYGVKRESIVVYVNEKKNALIIYPHGDHHDPI 426
QY 363 DEHKPWGICHSHSNYELFKPEEGYAKKEGKNVYTGELTNVYVLLKNSTFNQNTFLANG 422
Db 427 DEHKPWGICHSHSNYELFKPEEGYAKKEGKNVYTGELTNVYVLLKNSTFNQNTFLANG 486
QY 423 QKRVSFSPPELEKKGILNMLVKLITPDGKVLKESVKGVEGEGVGNIANFELDQYLPQ 482
Db 487 QKRVSFSPPELEKKGILNMLVKLITPDGKVLKESVKGVEGEGVGNIANFELDQYLPQ 546
QY 483 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDA 542
Db 547 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDA 606
QY 543 LVRVDFEFGHGNAYLNNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIV 602
Db 607 LVRVDFEFGHGNAYLNNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIV 666
QY 603 EYPTLEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTSEKVEKELSETGNSTNS 662
Db 667 EYPTLEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTSEKVEKELSETGNSTNS 726
QY 663 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQ 722
Db 727 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQ 786
QY 723 NGENKPSNGKVSIGTGTVENQPTENKPADSLPEAPNPKVPKPNSTNGMLNPEGVNSDP 846
Db 787 NGENKPSNGKVSIGTGTVENQPTENKPADSLPEAPNPKVPKPNSTNGMLNPEGVNSDP 846
QY 783 MLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVIRKKNL 837
Db 847 MLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVIRKKNL 901


```

RESULT 9
US-09-884-465A-372
; Sequence 372, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Carland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 372
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

      Query Match      73.2%      Score 3218:  DB 9:  Length 906:
      Best Local Similarity 73.3%; Pred. No. 6.3e-183;
      Matches 656; Conservative 35; Mismatches 84; Indels 120; Gaps 12;

QY  25  SQSSQKSENLTDPQVSQKEGIAQ-----QIVKIIDQG 58
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  45  AAQAAKEKGLTPPSTDHQDSGNTAEKAGAEALYNRVKAAKVPLDRMPYNLQYTVVEKNGS 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  59  YVTSHGPDHYHYNGKVPYDALFSELLMKDPNOLKADIVNEVKGYIIVKDGKYYVYL 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  105  LIITPHYDHYH--NIKFW---FDEGLYEAAPKGYSLD-----LLATYKYVVEP 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  119  KDAAHADNVRTKTDINRQKHVDKNEKNSNVAVARSGQRYTTNDGYYVFNPAIIIDTG 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  148  RNA--SDHVRK-----NKADQDSKFEDEKHEDEV---SETPHPESD-----EKEN 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  179  NAYIVPGGHGHHYTPKSDLSASELAAA----KAHAG----- 211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  188  HAGLNPASDNL-YKPSDTDETEEAEDTDEAIPGTPSIRQNAETLTGLKSSLLGT 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  212  -----KKNQPSOL-----SYSTSPSPSLINPGTSHEKHEE 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  247  KDNNTISAEDVDSLALLKESQAPIQGPQIGTLPNNLSATSPSLINPGTSHEKHEE 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  243  DGYGFANDRIIAEDESGFVMSGDHNHYFFKDLTEEQIKAAQKHLEEVKTSHNGLDSLS 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  307  DGYGFANDRIIAEDESGFVMSGSDSNHYFFKDLTEEQIKAAQKHLEEVKTSHNGLDSLS 366
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  303  SHEQDYPSNAKEMKDLDPKTEEKTAGTMKQYGVKRESIVVNEKNAIYPHGDHHDADI 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  367  SHEQDYPSNAKEMKDLDPKTEEKTAGTMKQYGVKRESIVVNEKNAIYPHGDHHDADI 426
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  363  DEHKPVGIGHSHSNYELFKPEEGVAKKEGNKVTGEBELTNVNLKKNSTNNQNTFLANG 422
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  427  DEHKPVGIGHSHSNYELFKPEEGVAKKEGNKVTGEBELTNVNLKKNSTNNQNTFLANG 486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  423  QKRVSFSPPELEKKGKLGINMLVLITPDGKLVLEKVSQKVFGEVGNNTANFELDOPYLPQG 482
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  487  QKRVSFSPPELEKKGKLGINMLVLITPDGKLVLEKVSQKVFGEVGNNTANFELDOPYLPQG 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  483  TFKYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLRVNPQFAVPKGTDA 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  547  TFKYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLRVNPQFAVPKGTDA 606
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  543  LVRVDFDEHGNAYLENNYKVGEEKLPIDKLNQGTTRTAGNKIPVTFWANAYLQDNQSYIV 602
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 607 LVRVDFEHGNAYLENNYKVGEIKLPIPKLNQCTTRTAGNKIPVTMANAYLDNQSTYIV 666
Qy 603 EVPILEXENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNST 662
Db 667 EVPILEXENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNST 726
Qy 663 LEEVPTVDPVQEVAKFAESYGGKKLENVLNFMQGTIELYLPSPGEVTKKNMADFTGEAPQG 722
Db 727 LEEVPTVDPVQEVAKFAESYGGKKLENVLNFMQGTIELYLPSPGEVTKKNMADFTGEAPQG 786
Qy 723 NGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPVKPNSTONGMLNPGNVGSDP 782
Db 787 NGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPVKPNSTONGMLNPGNVGSDP 846
Qy 783 MLDSALEEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVTKKNL 837
Db 847 MLDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVTKKNL 901

RESULT 10
US-09-884-465A-334
; Sequence 334, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 334
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-334

```

[illegible]

APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 361
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-361

Query Match 72.6%; Score 3193.5; DB 9; Length 633;
Best Local Similarity 98.4%; Pred. No. 1.1e-181;
Matches 614; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
1:
QY 215 QPSQLSYS-STPSPSLPINPGTSHKEHEEDGFGDANRIIADESGFVMSHGDHNNHYFFK 273
DB 5 QPTLPNNSLATPSPSLPINPGTSHKEHEEDGFGDANRIIADESGFVMSHGDHNNHYFFK 64
QY 274 KDLTEEQIAAKAKHLEEVKTSNGLDLSLSHEODYPSNAKEMKDLCKIEEKIAGIMKQY 333
DB 65 KDLTEEQIAAKAKHLEEVKTSNGLDLSLSHEODYPSNAKEMKDLCKIEEKIAGIMKQY 124
QY 334 GVKRESIVVWKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 393
DB 125 GVKRESIVVWKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 184
QY 394 VYTGELTNVNLKNSFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 453
DB 185 VYTGELTNVNLKNSFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 244
QY 454 LEKVGKVFEGVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 513
DB 245 LEKVGKVFEGVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 304
QY 514 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVGKILPIPKLN 573
DB 305 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVGKILPIPKLN 364
QY 574 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNOTDKPSILPQFKRNKAQENSKL 633
DB 365 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNOTDKPSILPQFKRNKAQENSKL 424
QY 634 DEKVEEPTSEKVEKEKLSGTNSTLSNLTLEVPVDPVQEKVAKFAESYGMKLENVLEN 693
DB 425 DEKVEEPTSEKVEKEKLSGTNSTLSNLTLEVPVDPVQEKVAKFAESYGMKLENVLEN 484
QY 694 MDGTIELYLPSPGEVKKMAADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLP 753
DB 485 MDGTIELYLPSPGEVKKMAADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLP 544
QY 754 EAPNEKPKPENSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 813
DB 545 EAPNEKPKPENSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 604
QY 814 SVIFNMDGTIELRLPSGEVKKNL 837
DB 605 SVIFNMDGTIELRLPSGEVKKNL 628

RESULT 15
US-09-884-465A-362
Sequence 362, Application US/09884465A
Publication No US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 362
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-362

Query Match 72.6%; Score 3193.5; DB 9; Length 633;
Best Local Similarity 98.4%; Pred. No. 1.1e-181;
Matches 614; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
QY 215 QPSQLSYS-STPSPSLPINPGTSHKEHEEDGFGDANRIIADESGFVMSHGDHNNHYFFK 273
DB 5 QPTLPNNSLATPSPSLPINPGTSHKEHEEDGFGDANRIIADESGFVMSHGDHNNHYFFK 64
QY 274 KDLTEEQIAAKAKHLEEVKTSNGLDLSLSHEODYPSNAKEMKDLCKIEEKIAGIMKQY 333
DB 65 KDLTEEQIAAKAKHLEEVKTSNGLDLSLSHEODYPSNAKEMKDLCKIEEKIAGIMKQY 124
QY 334 GVKRESIVVWKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 393
DB 125 GVKRESIVVWKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 184
QY 394 VYTGELTNVNLKNSFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 453
DB 185 VYTGELTNVNLKNSFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 244
QY 454 LEKVGKVFEGVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 513
DB 245 LEKVGKVFEGVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 304
QY 514 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVGKILPIPKLN 573
DB 305 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVGKILPIPKLN 364
QY 574 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNOTDKPSILPQFKRNKAQENSKL 633
DB 365 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNOTDKPSILPQFKRNKAQENSKL 424
QY 634 DEKVEEPTSEKVEKEKLSGTNSTLSNLTLEVPVDPVQEKVAKFAESYGMKLENVLEN 693
DB 425 DEKVEEPTSEKVEKEKLSGTNSTLSNLTLEVPVDPVQEKVAKFAESYGMKLENVLEN 484
QY 694 MDGTIELYLPSPGEVKKMAADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLP 753
DB 485 MDGTIELYLPSPGEVKKMAADFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLP 544
QY 754 EAPNEKPKPENSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 813
DB 545 EAPNEKPKPENSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 604
QY 814 SVIFNMDGTIELRLPSGEVKKNL 837
DB 605 SVIFNMDGTIELRLPSGEVKKNL 628

Search completed: May 13, 2003, 13:59:01
Job time : 28.7625 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:49:42 ; Search time 43.5434 Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAYALNQHRSQENKDNREVS.....IELRLPSGEVYKKNLSDFIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5315	100.0	1019	21	AA12722
2	5315	100.0	1019	21	AA12751
3	5315	100.0	1019	23	AAU84021
4	5315	100.0	1039	21	AA12715
5	5315	100.0	1039	23	AAU75932
6	5297	99.7	1019	21	AA12750
7	5291	99.5	1019	21	AA12748
8	5291	99.5	1019	21	AA12749
9	5291	99.5	1019	21	AA12752
10	5291	99.5	1019	21	AA12753

11	4325	81.4	1238	23	AAU84056	S. pneumoniae deri
12	4276.5	80.5	840	21	AA12721	Streptococcus pneu
13	4276.5	80.5	840	23	AAU76151	Streptococcus pneu
14	4263.5	80.2	1378	23	AAU84053	S. pneumoniae deri
15	4261.5	80.2	1365	23	AAU84057	S. pneumoniae deri
16	4244	79.8	1152	23	AAU84054	S. pneumoniae deri
17	4196	78.9	807	23	AAU83997	Truncated variant
18	4090	77.0	780	21	AA12744	Streptococcus pneu
19	3559	67.0	679	23	AAU84092	Truncated variant
20	3384	63.7	907	23	AAU84061	S. pneumoniae deri
21	3381	63.6	644	23	AAU83996	Truncated variant
22	3375	63.5	907	23	AAU84065	S. pneumoniae deri
23	3366	63.3	907	23	AAU84063	S. pneumoniae deri
24	3362	63.3	901	23	AAU84067	S. pneumoniae deri
25	3345.5	62.9	902	23	AAU84076	S. pneumoniae deri
26	3338	62.8	901	23	AAU84068	S. pneumoniae deri
27	3329	62.6	901	23	AAU84072	S. pneumoniae deri
28	3329	62.6	901	23	AAU84074	S. pneumoniae deri
29	3327.5	62.6	902	23	AAU84078	S. pneumoniae deri
30	3325	62.5	895	23	AAU84070	S. pneumoniae deri
31	3323.5	62.5	896	23	AAU84080	S. pneumoniae deri
32	3306	62.2	632	23	AAU83998	Truncated variant
33	3297	62.0	632	23	AAU84002	Truncated variant
34	3297	62.0	632	23	AAU84006	Truncated variant
35	3297	62.0	632	23	AAU84013	Truncated variant
36	3297	62.0	632	23	AAU84014	Truncated variant
37	3288.5	61.9	896	23	AAU84077	S. pneumoniae deri
38	3288	61.9	632	23	AAU84003	Truncated variant
39	3288	61.9	632	23	AAU84004	Truncated variant
40	3288	61.9	632	23	AAU84009	Truncated variant
41	3288	61.9	632	23	AAU84012	Truncated variant
42	3288	61.9	901	23	AAU84066	S. pneumoniae deri
43	3288	61.9	907	23	AAU84060	S. pneumoniae deri
44	3288	61.9	907	23	AAU84064	S. pneumoniae deri
45	3287	61.8	632	23	AAU84015	Truncated variant

ALIGNMENTS

RESULT 1
AA12722
ID AA12722 standard: Protein; 1019 AA.
XX AA12722;
XX
XX
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.

XX
XX WO200039299-A2.
XX
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hanel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX

Db 61 TSHGDHYHYNGKVPYDALEFSEELLMKDPNYQLKDDADIVNEVKGGYIIKVDGKYYVYVLKD 120

```
QY 121 AAHADNVRTKDEINRQKQEHVKDNKNSVNAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNKNSVNAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
QY 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
Db 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
QY 241 PANKSENLOSLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIPYSK 300
Db 241 PANKSENLOSLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIPYSK 300
QY 301 LSALEKIAKRWPISTGSTVSTNAKPNNEVSSGLSLNPNSSLTTSKELSSASDGYIFN 360
Db 301 LSALEKIAKRWPISTGSTVSTNAKPNNEVSSGLSLNPNSSLTTSKELSSASDGYIFN 360
QY 361 PKDIVETATAYIVRGDHFHYIPKSNQIGQPTLPNNSLATPSLPINPGTSHEKHEED 420
Db 361 PKDIVETATAYIVRGDHFHYIPKSNQIGQPTLPNNSLATPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSEGFVMSHGDNHYFFKKDLTEQIKAAQKHLEEVKTSHNGLDLSL 480
Db 421 GYGFDANRIIAEDSEGFVMSHGDNHYFFKKDLTEQIKAAQKHLEEVKTSHNGLDLSL 480
QY 481 HEODYPGNAKEMKDLKKTEETAGTMKQYGVKRESIVVKNKKNATIIYPHGHHADPTD 540
Db 481 HEODYPGNAKEMKDLKKTEETAGTMKQYGVKRESIVVKNKKNATIIYPHGHHADPTD 540
QY 541 EHKPVGIGHSHSYELFKPEEGVAKKEGKNVYTGESLTNVNLLKNSTNNQFTLANQ 600
Db 541 EHKPVGIGHSHSYELFKPEEGVAKKEGKNVYTGESLTNVNLLKNSTNNQFTLANQ 600
QY 601 KRVSFSPPELEKLGINMLVLIITPDGKVLKSVKQVGEVGNITANFELDQPLPGOT 660
Db 601 KRVSFSPPELEKLGINMLVLIITPDGKVLKSVKQVGEVGNITANFELDQPLPGOT 660
QY 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKNASQTIIFYPFHAGDTYLRVNPQFAVPKGTAL 720
Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKNASQTIIFYPFHAGDTYLRVNPQFAVPKGTAL 720
QY 721 VRVDFEFGHAYLNNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
Db 721 VRVDFEFGHAYLNNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
QY 781 VPILERENOTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKLSSETGNSNSTL 840
Db 781 VPILERENOTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKLSSETGNSNSTL 840
QY 841 EEPVTVDPVQEVKAPAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPGN 900
Db 841 EEPVTVDPVQEVKAPAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPGN 900
QY 901 GENKPSGKSVGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPSGNGSDPM 960
Db 901 GENKPSGKSVGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPSGNGSDPM 960
QY 961 LDPALIEEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 961 LDPALIEEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019

RESULT 3
AAU84021
ID AAU84021 standard; Peptide; 1019 AA.
XX
AC AAU84021;
XX
DT 08-MAY-2002 (first entry)
DE Truncated variant of S. pneumoniae BVH-3, BVH-3M.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
```

```
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX
PS Example 1; Page : 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation.
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1019 AA;
```

```
Query Match 100.0%; Score 5315; DB 23; Length 1019;
Best Local Similarity 100.0%; Pred. No 1.3e-314;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAYALNQHRSGENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIOAQEIIVKITDQGYV 60
Db 1 CAYALNQHRSGENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIOAQEIIVKITDQGYV 60
```

```
QY 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVNEVKGYIIVKDGKYYVLKD 120
Db 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVNEVKGYIIVKDGKYYVLKD 120
```

```
QY 121 AAHADNVRTKDEINRQKQEHVKDNKNSVNAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNKNSVNAVARSQGRYTTNDGYVFNPAIDIEDTGN 180
```

```
QY 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
Db 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
```

```
QY 241 PANKSENLOSLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIPYSK 300
Db 241 PANKSENLOSLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIPYSK 300
```

```

QY 301 LSALEEKIARVPIISGTSVSTNAKNEVVSLSGLSSNPSSLTTSKELSSASDGIYFN 360
Db 301 LSALEEKIARVPIISGTSVSTNAKNEVVSLSGLSSNPSSLTTSKELSSASDGIYFN 360
QY 361 PKDIVETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHEKHEED 420
Db 361 PKDIVETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
Db 421 GYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
QY 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVWKEKNALIIYPHGDHHDADPID 540
Db 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVWKEKNALIIYPHGDHHDADPID 540
QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKKGKQVYTGEBELTNVNLKSTFNQNFYLANGQ 600
Db 541 EHKPVGIGHSHSNYELFKPEEGVAKKKGKQVYTGEBELTNVNLKSTFNQNFYLANGQ 600
QY 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
Db 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
QY 661 FKYTITASKDYPEVSDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDAL 720
Db 661 FKYTITASKDYPEVSDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDAL 720
QY 721 VRVDFEFGNAYLNNYKVGKIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNOSTYIYE 780
Db 721 VRVDFEFGNAYLNNYKVGKIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNOSTYIYE 780
QY 781 VPILKENQTKPSILFOFKRNKAQENSKLDEKVEEPTSEKVEKSEKLSGTGNSNSTL 840
Db 781 VPILKENQTKPSILFOFKRNKAQENSKLDEKVEEPTSEKVEKSEKLSGTGNSNSTL 840
QY 841 EBPVTVDPQVKAKFAESYGMKLENVLFNNMDGTIELYLPSEGEVKKNNADFTGEAPQGN 900
Db 841 EBPVTVDPQVKAKFAESYGMKLENVLFNNMDGTIELYLPSEGEVKKNNADFTGEAPQGN 900
QY 901 GENKPESENGKYSTGTENQNTENKPADSLPEAPNEKPVKPENSTDNQMLNPEGVSGDPM 960
Db 901 GENKPESENGKYSTGTENQNTENKPADSLPEAPNEKPVKPENSTDNQMLNPEGVSGDPM 960
QY 961 LDPALAEAPADVPQVKLEKFTASYGGLDLSVIFNMDGTIELRLPSGEVKKNLSDFIA 1019
Db 961 LDPALAEAPADVPQVKLEKFTASYGGLDLSVIFNMDGTIELRLPSGEVKKNLSDFIA 1019

RESULT 4
AAB12715
ID AAB12715 standard; Protein: 1039 AA.
XX
AC AAB12715;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
PD
PF 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.

```

```

XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI: 2000-452397/39.
DR N-PSDB: AAA65730.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 2: 106pp; English.
XX
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
have bactericidal activity. The nucleic acids, encoding the protein
antigens, may be used for the recombinant production of the proteins
they encode. The protein antigens may then be used as vaccines for the
prevention and treatment of Streptococcal infections in mammals
(especially humans) which result in, e.g. meningitis, otitis media,
bacteraemia and/or pneumonia. The present sequence represents the
S. pneumoniae BVH-3 protein antigen.
XX
SQ Sequence 1039 AA;

```

```

Query Match 100.0%; Score 5315; DB 21; Length 1039;
Best Local Similarity 100.0%; Pred. No. 1.3e-314;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAYALNHRSEKNDNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db 21 CAYALNHRSEKNDNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGV 80
QY 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEVKGGYIIKVDGKYVYLKD 120
Db 81 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEVKGGYIIKVDGKYVYLKD 140
QY 121 AAHADNVRTDEINRQKQEHVKDNEKNSNVAVARSGRYTNDGVFNPADIIEDTGA 180
Db 141 AAHADNVRTDEINRQKQEHVKDNEKNSNVAVARSGRYTNDGVFNPADIIEDTGA 200
QY 181 YIVPHGHHYIYIPKSDLSASELAAAHLAGKNQSPQSLSYSTASDNNTQSVAGKSTK 240
Db 201 YIVPHGHHYIYIPKSDLSASELAAAHLAGKNQSPQSLSYSTASDNNTQSVAGKSTK 260
QY 241 PANKSENQLSKLLEYDPSAQRYSQSDGLVFPDAKIIISRTPNGVAIPHGDHYHFIYPSK 300
Db 261 PANKSENQLSKLLEYDPSAQRYSQSDGLVFPDAKIIISRTPNGVAIPHGDHYHFIYPSK 320
QY 301 LSALEEKIARVPIISGTSVSTNAKNEVVSLSGLSSNPSSLTTSKELSSASDGIYFN 360
Db 321 LSALEEKIARVPIISGTSVSTNAKNEVVSLSGLSSNPSSLTTSKELSSASDGIYFN 380
QY 361 PKDIVETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHEKHEED 420
Db 381 PKDIVETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHEKHEED 440
QY 421 GYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
Db 441 GYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 500
QY 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVWKEKNALIIYPHGDHHDADPID 540
Db 501 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVWKEKNALIIYPHGDHHDADPID 560
QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKKGKQVYTGEBELTNVNLKSTFNQNFYLANGQ 600
Db 561 EHKPVGIGHSHSNYELFKPEEGVAKKKGKQVYTGEBELTNVNLKSTFNQNFYLANGQ 620
QY 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
Db 621 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 680

```


QY 661 FKTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
 Db 681 FKTIASKDYPEVSYDGTFTVPTSLAYKMASQTFIFPFHAGDTYLRVNPQFAVPKGTDAL 740
 QY 721 VRYVDFHGNAYLENNYKYGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
 Db 741 VRYVDFHGNAYLENNYKYGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 800
 QY 781 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPFPKTSKVEKLEKSETGNSTNSL 840
 Db 801 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPFPKTSKVEKLEKSETGNSTNSL 860
 QY 841 EYVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPQGN 900
 Db 861 EYVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPQGN 920
 QY 901 GENKPSGCKVSGTGVENOPTENKPADSLPEAPNEKPVKPEPSTNDGMLNPGNVGSDPM 960
 Db 921 GENKPSGCKVSGTGVENOPTENKPADSLPEAPNEKPVKPEPSTNDGMLNPGNVGSDPM 980
 QY 961 LDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
 Db 981 LDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1039

RESULT 5
 AAU75932
 ID AAU75932 standard; Protein: 1039 AA.
 XX
 AC AAU75932;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-3 protein version #1.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA0908.
 XX
 XX 20-JUN-2000; 2000US-212883P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI: 2002-122272/16.
 DR N-PSDB; ABK15101.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Example 1; Fig 6; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This is the amino acid sequence of
 CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
 CC peptides described in the method of the invention.
 XX
 XX Sequence 1039 AA;
 Query Match 100.0%; Score 5315; DB 23; Length 1039;
 Best Local Similarity 100.0%; Pred. No. 1.3e-314;
 Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CAYALNQHRSEKNNRVSYYDGSQSSCKSENLTPOOVSOKEGIAQEQIVIKITDQYV 60
 Db 21 CAYALNQHRSEKNNRVSYYDGSQSSCKSENLTPOOVSOKEGIAQEQIVIKITDQYV 80
 QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIITKVGKYYVYLKD 120
 Db 81 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIITKVGKYYVYLKD 140
 QY 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGYTTNDGYVFPNADIETDQNA 180
 Db 141 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGYTTNDGYVFPNADIETDQNA 200
 QY 181 YTVPHGGHYHYIPKSDLSASELAAKAHLAKGNQPSOLSYSSTASDNTOSVAKGSTSK 240
 Db 201 YTVPHGGHYHYIPKSDLSASELAAKAHLAKGNQPSOLSYSSTASDNTOSVAKGSTSK 260
 QY 241 PANKSENQLSKLLKELYDPSAQRYSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
 Db 261 PANKSENQLSKLLKELYDPSAQRYSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 320
 QY 301 LSALEEKARWVPISGTGSTVSTNAKPNEVSSLSGSLSSNPSSLTTSKELASDGYIFN 360
 Db 321 LSALEEKARWVPISGTGSTVSTNAKPNEVSSLSGSLSSNPSSLTTSKELASDGYIFN 380
 QY 361 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHEKHEED 420
 Db 381 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHEKHEED 440
 QY 421 GYGFDANRIIADESGFYNSHGDHNNHYFFKDLTEEQKAAQKHLKEEVKTSKSHGLDLSLS 480
 Db 441 GYGFDANRIIADESGFYNSHGDHNNHYFFKDLTEEQKAAQKHLKEEVKTSKSHGLDLSLS 500
 QY 481 HEODYPGNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIVPHGDHHDADPID 540
 Db 501 HEODYPGNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIVPHGDHHDADPID 560
 QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGBELTNVNVNLLKNSTFNQNTFLANGQ 600
 Db 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGBELTNVNVNLLKNSTFNQNTFLANGQ 620
 QY 601 KRVSFSFPELEKKGILNMLVKLITPDGKVLKSKVSGKVGEGVGNIANFELDQYPLPQGT 660
 Db 621 KRVSFSFPELEKKGILNMLVKLITPDGKVLKSKVSGKVGEGVGNIANFELDQYPLPQGT 680
 QY 661 PKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
 Db 681 PKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAL 740
 QY 721 VRYVDFHGNAYLENNYKYGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
 Db 741 VRYVDFHGNAYLENNYKYGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 800
 QY 781 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPFPKTSKVEKLEKSETGNSTNSL 840
 Db 801 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPFPKTSKVEKLEKSETGNSTNSL 860
 QY 841 EYVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPQGN 900

```

Db      861  EVPTVDPVQEKVAKFAESYGMKLENVFNNDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
Qy      901  GENKPSNGKYSTGTVENOPTENKPADSLPEAPNEKPVKPDENSTDMNLPENGVGSDPM 960
Db      921  GENKPSNGKYSTGTVENOPTENKPADSLPEAPNEKPVKPDENSTDMNLPENGVGSDPM 980
Qy      961  LDPALPEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
Db      981  LDPALPEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1039

RESULT 6
AAB12750
ID      AAB12750 standard; Protein: 1019 AA.
XX
AC      AAB12750;
XX
DT      21-NOV-2000 (first entry)
XX
DE      Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
XX
KW      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW      prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      otitis media; pneumonia; immunisation; bactericidal.
XX
OS      Streptococcus pneumoniae.
XX
PN      WO200039299-A2.
XX
PD      06-JUL-2000.
XX
PF      20-DEC-1999; 99WO-CA01218.
XX
PR      23-DEC-1998; 98US-O113800.
XX
PA      (BIOC-) BIOCHEM PHARMA INC.
XX
PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX      WPI: 2000-452397/39.
XX
PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT      otitis media, bacteraemia and/or pneumonia -
XX      Disclosure; Fig 11; 106pp; English.
XX
CC      The present invention describes nucleic acids (I) encoding protein
CC      antigens (II) from Streptococcus pneumoniae. The protein antigens
CC      have bactericidal activity. The nucleic acids, encoding the protein
CC      antigens, may be used for the recombinant production of the proteins
CC      they encode. The protein antigens may then be used as vaccines for the
CC      prevention and treatment of Streptococcal infections in mammals
CC      (especially humans) which result in, e.g. meningitis, otitis media,
CC      bacteraemia and/or pneumonia. The present sequence represents a
CC      S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ      Sequence 1019 AA;

Query Match          99.7%; Score 5297; DB 21; Length 1019;
Best Local Similarity 99.7%; Pred. No. 1.6e-313;
Matches 1016; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1  CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSQKREGIAEQIVIKITDQGVV 60
Db      1  CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSQKREGIAEQIVIKITDQGVV 60
Qy      61  TSHGDHYHYNGKVPYDALFSEELLKRDPNYQLKADIYNEVKGHHIIVKDGKYYVYKLD 120
Db      61  TSHGDHYHYNGKVPYDALFSEELLKRDPNYQLKADIYNEVKGHHIIVKDGKYYVYKLD 120
Qy      121  AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPAADIIEDTGN 180

```

```

Db      121  AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPAADIIEDTGN 180
Qy      181  YIVPHGHYHYIIPKSDLSASELAARAAKHLAKKMQPSQLSYSTASDNNTQSVAGSTSK 240
Db      181  YIVPHGHYHYIIPKSDLSASELAARAAKHLAKKMQPSQLSYSTASDNNTQSVAGSTSK 240
Qy      241  PANKSENLSLLKELYDSPSAQRYSESGLVFPDPAKIIISRTPNGVAIPHDGHHYHFTPIYSK 300
Db      241  PANKSENLSLLKELYDSPSAQRYSESGLVFPDPAKIIISRTPNGVAIPHDGHHYHFTPIYSK 300
Qy      301  LSALAEKIARWPIISGTGTVSTNAKPNNEVVSGLSSNPPSSLTTSKELSSASOGYIFN 360
Db      301  LSALAEKIARWPIISGTGTVSTNAKPNNEVVSGLSSNPPSSLTTSKELSSASOGYIFN 360
Qy      361  PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPLIPNGTSHKHEED 420
Db      361  PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPLIPNGTSHKHEED 420
Qy      421  GYGFDAIRIIAEDSEGFVMSHGDNHNYFFKKDLTEQIKAAQKHLEEVKTSNGLDLS 480
Db      421  GYGFDAIRIIAEDSEGFVMSHGDNHNYFFKKDLTEQIKAAQKHLEEVKTSNGLDLS 480
Qy      481  HEQDYPGNAKEMDKDKKIEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPID 540
Db      481  HEQDYPGNAKEMDKDKKIEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPID 540
Qy      541  EHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKNSTFNNOFTLANGQ 600
Db      541  EHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKNSTFNNOFTLANGQ 600
Qy      601  KRVSFSPPELEKRLGGINMLVKLITPDGKVLKESGVKFECEGVGNIANFELDQVLPQGT 660
Db      601  KRVSFSPPELEKRLGGINMLVKLITPDGKVLKESGVKFECEGVGNIANFELDQVLPQGT 660
Qy      661  FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 720
Db      661  FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 720
Qy      721  VRVDEPHGNAYLNNYKVGKIEKIPKLNQGTTRTAGNKIPVTENMANAYLDNQSTYIVE 780
Db      721  VRVDEPHGNAYLNNYKVGKIEKIPKLNQGTTRTAGNKIPVTENMANAYLDNQSTYIVE 780
Qy      781  VPILKEKNTQDKSILPQFKRKAQENSKLDEKVESEPKTSEKVEREKISETGNSNSTL 840
Db      781  VPILKEKNTQDKSILPQFKRKAQENSKLDEKVESEPKTSEKVEREKISETGNSNSTL 840
Qy      841  EEVPTVDPVQEKVAKFAESYGMKLENVFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Db      841  EEVPTVDPVQEKVAKFAESYGMKLENVFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Qy      901  GENKPSNGKYSTGTVENOPTENKPADSLPEAPNEKPVKPDENSTDMNLPENGVGSDPM 960
Db      901  GENKPSNGKYSTGTVENOPTENKPADSLPEAPNEKPVKPDENSTDMNLPENGVGSDPM 960
Qy      961  LDPALPEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
Db      961  LDPALPEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019

RESULT 7
AAB12748
ID      AAB12748 standard; Protein: 1019 AA.
XX
AC      AAB12748;
XX
DT      21-NOV-2000 (first entry)
XX
DE      Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
XX
KW      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW      prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      otitis media; pneumonia; immunisation; bactericidal.
XX

```

OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 11; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 99.5%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 3.7e-313;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKDNRRYVYDGSOSKSENLPDQVSQREGIAEQIIVIKITDQGV 60
DB 1 CAYALNQHRSQENKDNRRYVYDGSOSKSENLPDQVSQREGIAEQIIVIKITDQGV 60
QY 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIKVDGKYYVYLD 120
DB 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIKVDGKYYVYLD 120
QY 121 AAHADNVRTKDEINRQKQEHVDNEKVNNAVARSQGYTTNDGVFNPADIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKQEHVDNEKVNNAVARSQGYTTNDGVFNPADIIEDTGN 180
QY 181 YIVPHGHYHYTPKSDLSASELAALAHLAGNMOPSQLSYSTASDNNITQSVAKGSTSK 240
DB 181 YIVPHGHYHYTPKSDLSASELAALAHLAGNMOPSQLSYSTASDNNITQSVAKGSTSK 240
QY 241 PANKSENQLLKLKLYDSPAQYSESDGLVDPKAIISRTENGVAIPHGDIHYFIPYSK 300
DB 241 PANKSENQLLKLKLYDSPAQYSESDGLVDPKAIISRTENGVAIPHGDIHYFIPYSK 300
QY 301 LSALEKIAMRVPISGTSTVNTAKPNWSSLSLNSNPSSLTSTKSLSSASDGYIFN 360
DB 301 LSALEKIAMRVPISGTSTVNTAKPNWSSLSLNSNPSSLTSTKSLSSASDGYIFN 360
QY 361 PKDIVEETATAVIRHGDHFHYPKSNQGTQPTLPNNSLATPSPLINPGTSHKHEED 420
DB 361 PKDIVEETATAVIRHGDHFHYPKSNQGTQPTLPNNSLATPSPLINPGTSHKHEED 420
QY 421 GYGFDANRIIADESGFVMSHGDHNYFFPKDLTTEQIAAKQKHLVEEYKTSNGLDLSS 480
DB 421 GYGFDANRIIADESGFVMSHGDHNYFFPKDLTTEQIAAKQKHLVEEYKTSNGLDLSS 480
QY 481 HEQDYPGNAKEMKLDKTEEKIAGIMKOYGVKRESIVVNEKKNAILIYPHGDHHPADPID 540
DB 481 HEQDYPGNAKEMKLDKTEEKIAGIMKOYGVKRESIVVNEKKNAILIYPHGDHHPADPID 540

QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTQELTNVYNLLKNSTFNQNTFLANGQ 600
DB 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTQELTNVYNLLKNSTFNQNTFLANGQ 600
QY 601 KRVSFSPPELEKELGGINMLVKLITPDCKVLEKVGSGKVGEGVGNIAFELDQYPLPGQT 660
DB 601 KRVSFSPPELEKELGGINMLVKLITPDCKVLEKVGSGKVGEGVGNIAFELDQYPLPGQT 660
QY 661 PKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAVPGKGTAL 720
DB 661 PKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPHAGDTYLRVNPQFAVPGKGTAL 720
QY 721 VVVFDEFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIVE 780
DB 721 VVVFDEFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIVE 780
QY 781 VPILKEKNTQKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTSNSTL 840
DB 781 VPILKEKNTQKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTSNSTL 840
QY 841 EEVPTVDVQEVKAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKNMADFTGEAPOGN 900
DB 841 EEVPTVDVQEVKAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKNMADFTGEAPOGN 900
QY 901 GENKPESENGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMGLNPEGNVGSDDPM 960
DB 901 GENKPESENGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMGLNPEGNVGSDDPM 960
QY 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMGTIELRLPSGVEIKKNSDFTA 1019
DB 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMGTIELRLPSGVEIKKNSDFTA 1019

RESULT 8
AAB12749
ID AAB12749 standard; Protein; 1019 AA.
XX AAB12749;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
XX WO200039299-A2.
PD 06-JUL-2000.
PF 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Disclosure; Fig 11; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match	99.5%;	Score 5291;	DB 21;	Length 1019;
Best Local Similarity	99.5%;	Pred. No. 3.7e-313;		
Matches 1014;	Conservative	2;	Mismatches 3;	Indels 0;
Gaps	0;			
QY	1	CAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGVV	60	
Db	1	CAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGVV	60	
QY	61	TSBGDHYHYNGKVPYDALFSEELLMKDPNYOLKADIVNEVKGYYIKVDGKYVYVYLD	120	
Db	61	TSBGDHYHYNGKVPYDALFSEELLMKDPNYOLKADIVNEVKGYYIKVDGKYVYVYLD	120	
QY	121	AAHADNVRTKDEINRQKHVDNEKVNNAVARSQGRYTTNDGVFNPAIDIEDTGN	180	
Db	121	AAHADNVRTKDEINRQKHVDNEKVNNAVARSQGRYTTNDGVFNPAIDIEDTGN	180	
QY	181	YIVPHGHHYHYIPKSDLSASELAHAGKAGNMOPSQLSYSTASDNNQTSQVAGSTSK	240	
Db	181	YIVPHGHHYHYIPKSDLSASELAHAGKAGNMOPSQLSYSTASDNNQTSQVAGSTSK	240	
QY	241	PANKSENLOQLKELYDSPAQSYSDGLVDPKAIISRTPNGVAIPHGDIHYHFIPIYSK	300	
Db	241	PANKSENLOQLKELYDSPAQSYSDGLVDPKAIISRTPNGVAIPHGDIHYHFIPIYSK	300	
QY	301	LSALEEKIARMVPISTGTVSTNAKPNVSVSLSSNPSSLTTSKELSSASDGIYFN	360	
Db	301	LSALEEKIARMVPISTGTVSTNAKPNVSVSLSSNPSSLTTSKELSSASDGIYFN	360	
QY	361	PKDIVEATAYIVRGDHPHYIPKSNQICQPTLPNNSLATPSPSLPINSPTSGHKEED	420	
Db	361	PKDIVEATAYIVRGDHPHYIPKSNQICQPTLPNNSLATPSPSLPINSPTSGHKEED	420	
QY	421	GYGFANRIIADDESFGVMSHGHNHYFFKKDLTEPQIKAAQKHLEEVKTSNGLDLSLS	480	
Db	421	GYGFANRIIADDESFGVMSHGHNHYFFKKDLTEPQIKAAQKHLEEVKTSNGLDLSLS	480	
QY	481	HEQDYPGNAMKMDLKKIEKTAGIMQKGVKRESIVNKENAIYIPHGDIHYHFIPIYSK	540	
Db	481	HEQDYPGNAMKMDLKKIEKTAGIMQKGVKRESIVNKENAIYIPHGDIHYHFIPIYSK	540	
QY	541	EHPKVGIGHSNHYELFKPEGVAKKGNKVVYTGEEELTNVNNLLKSTFNQNTLANGQ	600	
Db	541	EHPKVGIGHSNHYELFKPEGVAKKGNKVVYTGEEELTNVNNLLKSTFNQNTLANGQ	600	
QY	601	KRVSFSPPELEKKGILNMLVKLITPDGKLVLEKVSQKGVGEGVGNIANFELDQYLPQGT	660	
Db	601	KRVSFSPPELEKKGILNMLVKLITPDGKLVLEKVSQKGVGEGVGNIANFELDQYLPQGT	660	
QY	661	FKYTIASKDPEVSYDGTFTVPSLAKMASQITFYFPHAGDYLVRNPOFAPVPGKTGAL	720	
Db	661	FKYTIASKDPEVSYDGTFTVPSLAKMASQITFYFPHAGDYLVRNPOFAPVPGKTGAL	720	
QY	721	VRVDEFHGNAYLNNYKVGIEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE	780	
Db	721	VRVDEFHGNAYLNNYKVGIEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE	780	
QY	781	VPILEKENQTDKPSILPQFRKNAQENSKLDEKVEEPTSEKYEKELSETGNTSNTSL	840	
Db	781	VPILEKENQTDKPSILPQFRKNAQENSKLDEKVEEPTSEKYEKELSETGNTSNTSL	840	
QY	841	EEVPTVPQVQKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQGN	900	
Db	841	EEVPTVPQVQKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQGN	900	
QY	901	GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNSTDNGMLNPEGNVGSDDPM	960	
Db	901	GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNSTDNGMLNPEGNVGSDDPM	960	

Db 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNSTDNGMLNPEGNVGSDDPM 960
QY 961 LDPALPEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVYIKKNLSDFIA 1019
Db 961 LDPALPEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVYIKKNLSDFIA 1019

RESULT 9	
AAB12752	
ID AAB12752 standard; Protein; 1019 AA.	
XX AAB12752;	
DT 21-NOV-2000 (first entry)	
XX Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.	
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;	
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;	
KW otitis media; pneumonia; immunisation; bactericidal.	
XX Streptococcus pneumoniae.	
OS WO2000039299-A2.	
PN 06-JUL-2000.	
PD 20-DEC-1999; 99WO-CA01218.	
PF 23-DEC-1998; 98US-0113800.	
PR (BIOC-) BIOCHEM PHARMA INC.	
PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;	
PI WPI: 2000-452397/39.	
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,	
PT otitis media, bacteraemia and/or pneumonia	
PS Disclosure: Fig 11; 106pp; English.	
CC The present invention describes nucleic acids (I) encoding protein	
CC antigens (II) from Streptococcus pneumoniae. The protein antigens	
CC have bactericidal activity. The nucleic acids, encoding the protein	
CC antigens, may be used for the recombinant production of the proteins	
CC they encode. The protein antigens may then be used as vaccines for the	
CC prevention and treatment of Streptococcal infections in mammals	
CC (especially humans) which result in, e.g. meningitis, otitis media,	
CC bacteraemia and/or pneumonia. The present sequence represents a	
CC S. pneumoniae BVH-3 protein antigen, from the present invention.	
XX Sequence 1019 AA;	

Query Match	99.5%;	Score 5291;	DB 21;	Length 1019;
Best Local Similarity	99.6%;	Pred. No. 3.7e-313;		
Matches 1015;	Conservative	0;	Mismatches 4;	Indels 0;
Gaps	0;			
QY	1	CAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGVV	60	
Db	1	CAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGVV	60	
QY	61	TSBGDHYHYNGKVPYDALFSEELLMKDPNYOLKADIVNEVKGYYIKVDGKYVYVYLD	120	
Db	61	TSBGDHYHYNGKVPYDALFSEELLMKDPNYOLKADIVNEVKGYYIKVDGKYVYVYLD	120	
QY	121	AAHADNVRTKDEINRQKHVDNEKVNNAVARSQGRYTTNDGVFNPAIDIEDTGN	180	
Db	121	AAHADNVRTKDEINRQKHVDNEKVNNAVARSQGRYTTNDGVFNPAIDIEDTGN	180	
QY	181	YIVPHGHHYHYIPKSDLSASELAHAGKAGNMOPSQLSYSTASDNNQTSQVAGSTSK	240	
Db	181	YIVPHGHHYHYIPKSDLSASELAHAGKAGNMOPSQLSYSTASDNNQTSQVAGSTSK	240	

```
QY 241 PANKSENLOSILKELYDPSAORYSESDGLVDPDAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
DB 241 PANKSENLOSILKELYDPSAORYSESDGLVDPDAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
QY 301 LSALKEEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTSKELSSASDGYIFN 360
DB 301 LSALKEEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTSKELSSASDGYIFN 360
QY 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
DB 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 480
DB 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 480
QY 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNKKNAIYIPHGDDHHPID 540
DB 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNKKNAIYIPHGDDHHPID 540
QY 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELTNVNNLKNSTFNQNTFLANGQ 600
DB 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELTNVNNLKNSTFNQNTFLANGQ 600
QY 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVGEGVGNIANFELDQVLPQGT 660
DB 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVGEGVGNIANFELDQVLPQGT 660
QY 661 FRYTTASKDYPVSDGTFTVPTSLAYKMASQTIIPYFHAGDTYLRVNPQFAVPGKTDAL 720
DB 661 FRYTTASKDYPVSDGTFTVPTSLAYKMASQTIIPYFHAGDTYLRVNPQFAVPGKTDAL 720
QY 721 VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIIVE 780
DB 721 VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIIVE 780
QY 781 VPILKENQTDKPSILPOFKRKAAGENSCLDEKVEEPTSEKVEKEKLSSETGNSNSTL 840
DB 781 VPILKENQTDKPSILPOFKRKAAGENSCLDEKVEEPTSEKVEKEKLSSETGNSNSTL 840
QY 841 EBPVTVDPQEVKAVFAESYGMKLENVLFNMDGTIELYLPSEVTKKNMADFTGEAPGN 900
DB 841 EBPVTVDPQEVKAVFAESYGMKLENVLFNMDGTIELYLPSEVTKKNMADFTGEAPGN 900
QY 901 GENKPSNGKSVGTGVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPM 960
DB 901 GENKPSNGKSVGTGVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPM 960
QY 961 LDPALAEAPVDPQVEKLEKFTASGLGLDSVIFNMQGTIELRLPSGEVIRKKNLSDFTA 1019
DB 961 LDPALAEAPVDPQVEKLEKFTASGLGLDSVIFNMQGTIELRLPSGEVIRKKNLSDLIA 1019

RESULT 10
AAB12753
ID AAB12753 standard; Protein; 1019 AA.
XX
AC AAB12753;
XX
XX 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-3 protein antigen.
DE
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
XX 06-JUL-2000.
PD
```

```
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-Q113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Disclosure; Fig 11: 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
XX Sequence 1019 AA;
XX
Query Match 99.5%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 3.7e-313;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CAYALNHRSEKNDNRVSVYDGSQSSQSKSENITPDQVSKQEGIAEQIVKITDGYV 60
DB 1 CAYALNHRSEKNDNRVSVYDGSQSSQSKSENITPDQVSKQEGIAEQIVKITDGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKLDADIVNEVKGYYIKVDGYVYVVKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKLDADIVNEVKGYYIKVDGYVYVVKD 120
QY 121 AAHADNVRTKDEINRQKQEHVKONEKYNVNAVARSQRYTNDGYVFNADIEDTGN 180
DB 121 AAHADNVRTKDEINRQKQEHVKONEKYNVNAVARSQRYTNDGYVFNADIEDTGN 180
QY 181 YIVPHGGHYHYIPKSDLSASLAAKAGKKNMOPQSLSYSTASDNNTQSVAKGSTK 240
DB 181 YIVPHGGHYHYIPKSDLSASLAAKAGKKNMOPQSLSYSTASDNNTQSVAKGSTK 240
QY 241 PANKSENLOSILKELYDPSAORYSESDGLVDPDAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
DB 241 PANKSENLOSILKELYDPSAORYSESDGLVDPDAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
QY 301 LSALKEEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTSKELSSASDGYIFN 360
DB 301 LSALKEEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTSKELSSASDGYIFN 360
QY 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
DB 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 480
DB 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLS 480
QY 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNKKNAIYIPHGDDHHPID 540
DB 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNKKNAIYIPHGDDHHPID 540
QY 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELTNVNNLKNSTFNQNTFLANGQ 600
DB 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELTNVNNLKNSTFNQNTFLANGQ 600
QY 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVGEGVGNIANFELDQVLPQGT 660
```

Db 601 KRVSEFPPELEKKGILNMLVKLITPDGKLEKVGKVGEGVGNIANFELDQYLPQGT 560
Qy 661 FRYTTASKDYPEVSDGFTFTVTSLAYKMASOTIYPPHAGDTYLRVNPQFAVPKGTAL 720
Db 661 FRYTTASKDYPEVSDGFTFTVTSLAYKMASOTIYPPHAGDTYLRVNPQFAVPKGTAL 720
Qy 721 VRVDFEFGHAYLENNYKVGTEKLPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 780
Db 721 VRVDFEFGHAYLENNYKVGTEKLPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVE 780
Qy 781 VPILEKENOTDKPSLPQKRNKAQENSKLDKVEEPTSEKVEKEKSEKTSNSTL 840
Db 781 VPILEKENOTDKPSLPQKRNKAQENSKLDKVEEPTSEKVEKEKSEKTSNSTL 840
Qy 841 EEVPTVDVQEVKFAESYGMKLENLFNMDGTIELVPSGEVKKNAADFTGEAPOGN 900
Db 841 EEVPTVDVQEVKFAESYGMKLENLFNMDGTIELVPSGEVKKNAADFTGEAPOGN 900
Qy 901 GENKPSGKSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDPM 960
Db 901 GENKPSGKSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNVGSDPM 960
Qy 961 LDPALAEAPAVDPQVKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNLSDFA 1019
Db 961 LDPALAEAPAVDPQVKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNLSDFA 1019

RESULT 11
AAU84056
ID AAU84056 standard; Peptide; 1238 AA.
XX AAU84056;
XX
XX 08-MAY-2002 (first entry)
DE S. pneumoniae derived chimeric peptide, NEW29.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutein;
XX BVH-11-2.
XX
XX Streptococcus pneumoniae.
OS Synthetic.
XX
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX PI
XX WPI: 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia
XX
XX Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 1238 AA:
Qy
Query Match 81.4%; Score 4325; DB 23; Length 1238;
Best Local Similarity 70.7%; Pred. No. 2.6e-254; Indels 260; Gaps 12;
Matches 876; Conservative 39; Mismatches 64;
Qy 25 SQSSQKSEN--TPDQVSQKEGIAEQIVIKITDQGYVTSBGHGHYHYNGKVPYDALFSE 82
Db 16 SQRHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF---IPYEQMSEL 59
Qy 83 EL-----LKKDPNYQKDA-----EHVKD--NEKV-----DI 98
Db 60 EKRIARIPIRYRSNHVWPDSPRPEPSPQTPPEPSPQAPNPQAPNPIDEKLVKEA 119
Qy 99 VNEVKGVIKVDG-KYIVVLK-----AAHADNVRTKDEI-----133
Db 120 VRKVGDCGVFENCVSRYIPAKNLASAETAAGIDSKLAKQESLSHKLGAKKATDLPSSDREF 179
Qy 134 -----NRQKQ-----EHVKD--NEKV-----147
Db 180 YNKAYDULARITHODLNDKGRQVDFAELNDLLEKRLKQVSSDKVLVDLAFIAPTRHPE 239
Qy 148 -----NSNVAVARSGRYTTNDGVFNPADIIEDTGNAYIVPHGGHYHYIPKS 195
Db 240 RLGPENQAITYTDEIQVAKLAGKYTTEDGYIFDPDITSDGDAYVTPHMTSHWIKKD 299
Qy 196 DLSASELAAKAHL-----209
Db 300 SLSEARAAQAQAYAKEKGLTPPSTDHDSGNTAKGAEIYNRVKAARKVPLDRMPYNLQ 359
Qy 210 -----209
Db 360 YTVKNGSLIPIHYDHYHNKIFWFDEGLYEPKGYTLEDLATVYVYVHEHNERPHSD 419
Qy 210 -----AGKNMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLKLKELVDSPS 260
Db 420 NGFGNASDHVGNPQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLKLKELVDSPS 479
Qy 261 AORYSESDGLVDFPAKIIISRTPNGVAPHGDHGFIFDYKLSALEEKIARKVPISTGST 320
Db 480 AORYSESDGLVDFPAKIIISRTPNGVAPHGDHGFIFDYKLSALEEKIARKVPISTGST 539
Qy 321 VSTNAKPNEVVSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHF 380
Db 540 VSTNAKPNEVVSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHF 599
Qy 381 HYIPKSNQIGQFTLPNNSLATPSPSLINPGTSHKHEEDGYGFDANRIIAEDSEGFVMS 440
Db 600 HYIPKSNQIGQFTLPNNSLATPSPSLINPGTSHKHEEDGYGFDANRIIAEDSEGFVMS 659
Qy 441 HGDENHVFEEKDLTEQIKAAQKHEEVKTSNGLDLSLSHEEDYPCGNAKEMKDLKKIE 500
Db 660 HGDENHVFEEKDLTEQIKAAQKHEEVKTSNGLDLSLSHEEDYPCGNAKEMKDLKKIE 719
Qy 501 EKTAGIMKQYGVKRESIVVKNKNAIITYPHGDHHAADPIDEHKPVGIGHSHSNVELEFKPE 560
Db 720 EKTAGIMKQYGVKRESIVVKNKNAIITYPHGDHHAADPIDEHKPVGIGHSHSNVELEFKPE 779

QY 561 EGVAKREGKVVYTGEEELTNVNVLLKNSFNQNFLLANGOKRVSFPPPELEKKGILNML 620
DB 780 EGVAKREGKVVYTGEEELTNVNVLLKNSFNQNFLLANGOKRVSFPPPELEKKGILNML 839
QY 621 VKLITPDGKVLKSVKVFGEVGNANFELDPYLPQGTFFKTYTIAASKDYPEVSYDGTFT 680
DB 840 VKLITPDGKVLKSVKVFGEVGNANFELDPYLPQGTFFKTYTIAASKDYPEVSYDGTFT 899
QY 581 VPTS LAYKMASQIFVFFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHGNAYLENNYKVG 740
DB 900 VPTS LAYKMASQIFVFFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHGNAYLENNYKVG 959
QY 741 EIKLPIPKLNOGTRAGNKIPVTFEMANAYLDNQSYIIVEVPILEKENOTDKPSILPQEK 800
DB 960 EIKLPIPKLNOGTRAGNKIPVTFEMANAYLDNQSYIIVEVPILEKENOTDKPSILPQEK 1019
QY 801 RNKAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQEKVAKFAESY 860
DB 1020 RNKAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQEKVAKFAESY 1079
QY 861 GMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKSVGTGVENOP 920
DB 1080 GMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKSVGTGVENOP 1139
QY 921 TENKPADSLPEAPNEKPKPENSTNGMLNPEGNVSGSDPMLDPALEAPAVDPVQEKLEK 980
DB 1140 TENKPADSLPEAPNEKPKPENSTNGMLNPEGNVSGSDPMLDPALEAPAVDPVQEKLEK 1199
QY 981 FTASYGLGLDSVIFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKSVGTGVENOP 1019
DB 1200 FTASYGLGLDSVIFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKSVGTGVENOP 1238

RESULT 12

AA12721
ID AA12721 standard; Protein; 840 AA.
XX AC AA12721;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
DE DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX DR N-PSDB; AAA65738.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX Claim 18; Fig 19; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae SP63 BVH-3 protein antigen.
XX
SQ Sequence 840 AA;

Query Match 80.5%; Score 4276.5; DB 21; Length 840;

Best Local Similarity 82.3%; Pred. No. 1.4e-251;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHSQENKONNRVSYVDGSSQSKSENLPDQVSQKEGIAQCIVIKITDQGIY 60
DB 1 CAYALNQHSQENKONNRVSYVDGSSQSKSENLPDQVSQKEGIAQCIVIKITDQGIY 60

QY 61 TSHGDHYHYNGKVPYDALFSELLMKOPNYOLKADIVNEVKGYYIIKVQDGKYYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSELLMKOPNYOLKADIVNEVKGYYIIKVQDGKYYVYLKD 120

QY 121 AAHADNVRTKDEINQKQEHVKDNKVNNAVARSQRYTNDGYVFNPAIDIEDTGN 180
DB 121 AAHADNVRTKDEINQKQEHVKDNKVNNAVARSQRYTNDGYVFNPAIDIEDTGN 180

QY 181 YIVPHGGRVHYTPKSDLSASELAAKAHLAKGNMOPSQLSYSTASDNNTOSVANGSTK 240
DB 181 YIVPHGGRVHYTPKSDLSASELAAKAHLAKGNMOPSQLSYSTASDNNTOSVANGSTK 240

QY 241 PANKSENIOQLKELYDPSAQRYSQGLVFDPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
DB 241 PANKSENIOQLKELYDPSAQRYSQGLVFDPAKIIISRTPNGVAIPHGDHYHFIPYSK 300

QY 301 LSALEEKIARVYPISTGTSTVSTNAKPNEVYSSLSGNPSSLTTSKELSSASDGIYFN 360
DB 301 LSALEEKIARVYPISTGTSTVSTNAKPNEVYSSLSGNPSSLTTSKELSSASDGIYFN 360

QY 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
DB 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420

QY 421 GYGFDPANRTIAEDSGFVNSHGDHNYFFKDLTEEQKAAQKHLEEVKTSHGNDLSLS 480
DB 421 GYGFDPANRTIAEDSGFVNSHGDHNYFFKDLTEEQKAAQKHLEEVKTSHGNDLSLS 480

QY 481 HQQDYPGNAKEMKDLKDKIEEKIAGIMQYGVKRESIIVNKEKNAIYYPHGDHHAID 540
DB 481 HQQDYPGNAKEMKDLKDKIEEKIAGIMQYGVKRESIIVNKEKNAIYYPHGDHHAID 540

QY 541 EHKPVGIGHSNHYELFKPEGVAKKGNKVYTGEEELTNVNVLLKNSFNQNFLLANGQ 600
DB 541 EHKPVGIGHSNHYELFKPEGVAKKGNKVYTGEEELTNVNVLLKNSFNQNFLLANGQ 600

QY 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKSVKVFGEVGNANFELDPYLPQGT 660
DB 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKSVKVFGEVGNANFELDPYLPQGT 660

QY 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQIFVFFHAGDTYLRVNPQFAVPKGTDAL 720
DB 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQIFVFFHAGDTYLRVNPQFAVPKGTDAL 720

QY 721 VRVDFEFHGNAYLENNYKVGKILPIPKLNOGTRTAGNKIPVTFEMANAYLDNQSYI 780
DB 721 VRVDFEFHGNAYLENNYKVGKILPIPKLNOGTRTAGNKIPVTFEMANAYLDNQSYI 780

QY 781 VPILKENQTDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840
DB 781 VPILKENQTDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840

QY 841 BEVPTVDPVQEKVAKFAESYGNKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 900
DB 841 BEVPTVDPVQEKVAKFAESYGNKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 900

QY 664 BEVPTVDPVQEKVAKFAESYGNKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 723
DB 664 BEVPTVDPVQEKVAKFAESYGNKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 723

901	QY	GENKPSNGKVSTGTVENOPTENKNPADSLPEAPENKPKVPKPENSTDNGMLNPNCGVSDPM	960
724	Db	GENKPSNGKVSTGTVENOPTENKNPADSLPEAPENKPKVPKPENSTDNGMLNPNCGVSDPM	783
961	QY	LDPALSEAPAYDVBQVKLKFTASYGLGDSVFNNMDGTIELRPLSPGEVIKKNL	1014
784	Db	LDSALSEAPADPBQVKLKFTASYGLGDSVFNNMDGTIELRPLSPGEVIKKNL	837

RESULT 13

AAU76151
ID AAU76151 standard; Protein; 840 AA.

AA
AC
ADU76151.

[illegible]

xx Streptococcus pneumoniae BVH-3 protein version #2.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection.

Streptococcus pneumoniae.

XX PN WO200198334-A2.

27-DEC-2001.

XX
PF 19-JUN-2001; 2001WO-CA00908.XX
PR 20-JUN-2000: 2000US-212683P.XX
PA (SHTR-) SHTR BIOCHEM INC.

XX
PT Hamel J., Ouellet C., Charla C., Charland N., Martin D., Brodeur B;

XX
DR WPI: 2002-122272/16.

DK N-PSUB; ABR13103.
XX
PT New streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
PT

XX
PS
Example 1; Fig 10; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chemic sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This is the amino acid sequence of Streptococcus pneumoniae protein BVH-3, used to create the antigenic peptides described in the method of the invention.

XX
S0 Sequence 840 AA;

Query Match	80.5%	Score	4276.5	DB	23	Length	840
Best Local Similarity	82.3%	Pred. No.	1.4e-251				
Matches	835	Conservative	0	Mismatches	2	Indels	177
						Gaps	1

RESULT 14
AAU84053
ID AAU8
XX

AC AAU84053;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, NEW26.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 KW BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 PN WO200198334-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA0908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR
 XX
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX
 XX Example 1; Page -; 113pp; English.
 PS
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (III) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A.
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 1378 AA;
 Query Match 80.2%; Score 4263.5; DB 23; Length 1378;
 Best Local Similarity 53.7%; Pred. No. 1.7e-250;
 Matches 879; Conservative 38; Mismatches 60; Indels 403; Gaps 13;
 QY 25 SSSQSKSENL--TPDOVSQKEGIAQCIYKIKTDQGYVTSBGDHYHYNGKYPYDALFSE 82
 DB 17 SERHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF----IPYEQMSL 60
 QY 83 E-----LLMKDQPNYQ-----LKDA 96
 DB 61 EKRIARIPIRLYRNSHWPDSEQSPQSTPEPSPSLOPAPNPQPSNPIDEKLVKEA 120
 QY 97 DIYNEVKGHHIKVDG-KYIVYIKD-----AAHADNVRKDEI-----133

DB 121 --VRKVGQYVFEENGVSRYIPAKDLISAETAAGIDSKIAKQESLSHKLGAKKTOLPSSDR 178
 QY 134 -----NROKQ-----BHVKD--NEKV-----147
 DB 179 EFYNKAYDLLARIHQDLDNKGQVDFEVLNLLERLKGVSDDKVKLVDDILAFIAPIRH 238
 QY 148 -----NSNVAVARSQORYTNDGYVFNPAIDIEDTGNAYIVPHGGHYHIP 193
 DB 239 PERLQKPNQIITYTDDIETQVAKLAGKYTEDGYIFDPRDITSDGDAYVTPHMTSHWIK 298
 QY 194 KSDLSASSELAAKAH-----208
 DB 299 KDSLSEAEERAAQAQAKGKGLTPPTDHDGSGNTEAKGAEAIYNVKAACKVPLDRMYPN 358
 QY 209 -----208
 DB 359 LOYTVKNGSLIIPHYDHYHNKFEWFEDEGLYEAPKGYSLDLLATYKYVVEHNERPH 418
 QY 209 -----208
 DB 419 SDNGFGNASDHYRNKKAQODSKPDEDKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPS 478
 QY 209 -----208
 DB 479 TDTEETEBAEDTTDEAETPQVENSVINAKIADAEALKEKVTDPISIRQNAMEITLGLKSS 538
 QY 209 -----LAKNNQPSQLSVSYSTASNDNTOSVAKGSTS 239
 DB 539 LLLGTGKDNNTISAEVDSLLALKESQAPIQGNMQPSQLSVSYSTASNDNTOSVAKGSTS 598
 QY 240 KPANKSENLSLLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIYS 299
 DB 599 KPANKSENLSLLKELYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIYS 658
 QY 300 KLSALEEKIARMPVPSGCTSTVSTNAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIF 359
 DB 659 KLSALEEKIARMPVPSGCTSTVSTNAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIF 718
 QY 360 NPKDIVEETATAYIVRHGDHFHYIPKSNQIQGOTLPNNSLATPPSPSLPINPCTSEKHEE 419
 DB 719 NPKDIVEETATAYIVRHGDHFHYIPKSNQIQGOTLPNNSLATPPSPSLPINPCTSEKHEE 778
 QY 420 DOYGFDAIRIADDESFGVMSGDHNNHYFFKDLTEEQKAAQKHEEVKTSHGNDLSLS 479
 DB 779 DOYGFDAIRIADDESFGVMSGDHNNHYFFKDLTEEQKAAQKHEEVKTSHGNDLSLS 838
 QY 480 SHEQDYPGNAKEMKDLDKKIEKIAQIMKQYGVKRESIWNKEKNAIIVPHGDHHAADI 539
 DB 839 SHEQDYPGNAKEMKDLDKKIEKIAQIMKQYGVKRESIWNKEKNAIIVPHGDHHAADI 898
 QY 540 DEHKPVGISHSHSNYELFKPBGVAKKGNKYVTGEELTNVYNLLKNSTFNQNFLLANG 599
 DB 899 DEHKPVGISHSHSNYELFKPBGVAKKGNKYVTGEELTNVYNLLKNSTFNQNFLLANG 958
 QY 600 QKRVSEFSPPELEKKLGNLMLVKLITPDGKVLKYSKGVGEGVGNIANFELDQPYLPQG 659
 DB 959 QKRVSEFSPPELEKKLGNLMLVKLITPDGKVLKYSKGVGEGVGNIANFELDQPYLPQG 1018
 QY 660 TKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVEKGTDA 719
 DB 1019 TKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVEKGTDA 1078
 QY 720 LYRVDFEFGHGNAYLENNYKVGKIKPIPKLNGGTTTTRAGNKIPVTFMANAYLDNQSTIV 779
 DB 1079 LYRVDFEFGHGNAYLENNYKVGKIKPIPKLNGGTTTTRAGNKIPVTFMANAYLDNQSTIV 1138
 QY 780 EYPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPEKTSKVEKEKLSETGNTSNT 839
 DB 1139 EYPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPEKTSKVEKEKLSETGNTSNT 1198
 QY 840 LEEVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQG 899
 DB 1199 LEEVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQG 1258

Db 1020 SYDGTFTVPTSLAYKMASQTFVFFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYL 1079
QY 734 ENNYKVCIEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKP 793
Db 1080 ENNYKVCIEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKP 1139
QY 794 SILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTLNSTLEEVPTVDPVQEKV 853
Db 1140 SILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTLNSTLEEVPTVDPVQEKV 1199
QY 854 AKFAESYGMKLENVLFNMDGTIELYLPSEVIRKKNMADFTGEAPQNGENKPSSENGKSVST 913
Db 1200 AKFAESYGMKLENVLFNMDGTIELYLPSEVIRKKNMADFTGEAPQNGENKPSSENGKSVST 1259
QY 914 GTVENQPTENKPADSLPEAPNEKPVKPFENSTDNMGMLNPEGNVGSDDPMLDPALEEFPAVDP 973
Db 1260 GTVENQPTENKPADSLPEAPNEKPVKPFENSTDNMGMLNPEGNVGSDDPMLDPALEEFPAVDP 1319
QY 974 VQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSEVIRKKNLSDFIA 1019
Db 1320 VQEKLEKFTASYGLGLSDSVIFNMDGTIELRLPSEVIRKKNLSDFIA 1365

Search completed: May 13, 2003, 13:53:53
Job time : 52.5434 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:27 ; Search time 23.0524 seconds
(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315
Sequence: 1 CAVALNQHRSQENKDNRRVS.....IELRLPSGVEIKKNLSDFIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5297	99.7	1039	2 H95115	conserved hypotet
2	5291	99.5	1039	2 D97985	hypothetical prote
3	1292.5	24.3	855	2 D98004	histidine Motif-Co
4	1263.5	23.8	853	2 C97985	hypothetical prote
5	1261	23.7	828	2 E98004	hypothetical prote
6	1255	23.6	802	2 C95136	conserved domain p
7	1235.5	23.2	819	2 B95136	conserved domain p
8	1233	23.2	839	2 C95115	conserved hypotet
9	899.5	16.9	822	2 T46758	hypothetical 92.4K
10	324.5	6.1	182	2 F97985	hypothetical prote
11	238.5	4.5	2004	2 F95133	immunoglobulin A1
12	228	4.3	2485	1 H71621	serine/threonine-s
13	223	4.2	1963	2 B98002	IgA-specific metal
14	220	4.1	1701	2 A54498	major merozoite su
15	218	4.1	1873	2 T30944	surface protein pr
16	214	4.0	1939	2 T18372	repeat organellar
17	209	3.9	1701	2 A26868	major merozoite su
18	205.5	3.9	1185	2 A42404	collagen adhesin -
19	203.5	3.8	1849	2 C41859	IgA-specific metal
20	199	3.7	1726	1 SAZQOM	major merozoite su
21	196	3.7	1271	2 A45555	glutamate rich pro
22	196	3.7	1726	2 A45948	major merozoite su
23	193	3.6	3724	2 T18427	hypothetical prote
24	192	3.6	5005	2 F82884	hypothetical prote
25	190.5	3.6	4549	2 T20771	hypothetical prote
26	190.5	3.6	4667	2 T20774	hypothetical prote
27	189	3.6	1038	2 H90053	hypothetical prote
28	187.5	3.5	2144	2 A97942	metalloproteinase
29	187	3.5	1080	2 T43164	TactA protein - Li

ALIGNMENTS

RESULT 1

H95115

conserved hypothetical protein SPI004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Molecule type: DNA
A:Status: preliminary

A:Residues: 1-1039 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

A:Genetics:

A:Gene: SPI004

Query Match 99.7%; Score 5297; DB 2; Length 1039;

Best Local Similarity 99.7%; Pred. No 9e-254; Mismatches 0; Indels 0; Gaps 0;

Matches 1016; Conservative 0; Indels 0; Gaps 0;

Qy 1 CAVALNQHRSQENKDNRRVSVDGSSQSKSENLTPDQVSKQEGIAEQIVIKITDQGVV 60

Db 21 CAVALNQHRSQENKDNRRVSVDGSSQSKSENLTPDQVSKQEGIAEQIVIKITDQGVV 80

Qy 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGKGIKVDGKYYVYLKD 120

Db 81 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGKGIKVDGKYYVYLKD 140

Qy 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGVYFNPADIEDTGA 180

Db 141 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGVYFNPADIEDTGA 200

Qy 181 YIPPHGCHYHYIPKSDLSASLAAKHAHLAKGNQPSQLSYSTASDNTQSVAKGSTK 240

Db 201 YIPPHGCHYHYIPKSDLSASLAAKHAHLAKGNQPSQLSYSTASDNTQSVAKGSTK 260

Qy 241 PANKSENQLLKELYDPSAQRYSESDGLVFPDPAKTIISRTPNGVAIPHGDHVFHFFPYSK 300

Db 261 PANKSENQLLKELYDPSAQRYSESDGLVFPDPAKTIISRTPNGVAIPHGDHVFHFFPYSK 320

Qy 301 LSALEEKIARMPISGTGTVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGYIFN 360

Db 321 LSALEEKIARMPISGTGTVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGYIFN 380

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Qy 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 361 PKDIVERATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED 420

Db 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSLPIINPGTSHKHEED 440
QY 421 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSINGLDSLS 480
Db 441 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSINGLDSLS 500
QY 481 HEQDYPGNKEMKDLKIEEIKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPID 540
Db 501 HEQDYPGNKEMKDLKIEEIKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPID 560
QY 541 EHPVGIGSHSNYELFKPEEGVAKKEGKVTGTELTNVNLLKNSFNQNFPLANGQ 600
Db 561 EHPVGIGSHSNYELFKPEEGVAKKEGKVTGTELTNVNLLKNSFNQNFPLANGQ 620
QY 601 KRVSFPPELEKKGILNMLVKLITPDGKVLKESYGVGEGVGNIANFELDPYLPQGT 660
Db 621 KRVSFPPELEKKGILNMLVKLITPDGKVLKESYGVGEGVGNIANFELDPYLPQGT 680
QY 661 FKTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDAL 720
Db 681 FKTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDAL 740
QY 721 VRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNQSTIYVE 780
Db 741 VRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNQSTIYVE 800
QY 781 VPILKENDQKPSILPOFKRNKAQENSLDKVEPEPKTSEKVEKLSNLSNSTL 840
Db 801 VPILKENDQKPSILPOFKRNKAQENSLDKVEPEPKTSEKVEKLSNLSNSTL 860
QY 841 EEPVTPDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Db 861 EEPVTPDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
QY 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVGSDPM 960
Db 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVGSDPM 980
QY 961 LDPALAEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
Db 981 LDPALAEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIKKNLSDLIA 1039

RESULT 2
D97985
hypoetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174
C:Genetics:
A:Gene: phtE

Query Match 99.5%; Score 5291; DB 2: Length 1039;
Best Local Similarity 99.5%; Pred. No. 1.8e-253;
Matches 1014; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAVALNHRQENKDNRRSYVDSOSSKSENLTDPVQSKREGIOAEQIVKITDQGVY 60
Db 21 CAVALNHRQENKDNRRSYVDSOSSKSENLTDPVQSKREGIOAEQIVKITDQGVY 80
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPYQLKDADIVNEVGKGYIIKVDGKYVYLKD 120

Db 81 TSHGDHYHYNGKVPYDALFSEELLMKDPYQLKDADIVNEVGKGYIIKVDGKYVYLKD 140
QY 121 AAHADNVRTKDEINRQKQEHVKONEKYNVAVARSOGRYTTNDGYVFNPAADIEDTGN 180
Db 141 AAHADNVRTKDEINRQKQEHVKONEKYNVAVARSOGRYTTNDGYVFNPAADIEDTGN 200
QY 181 YIVPHGHHYHVIKPSOLASASAAKHAHLAKGNMOPQSOLSYSTASDNNTQSVAKGSTK 240
Db 201 YIVPHGHHYHVIKPSOLASASAAKHAHLAKGNMOPQSOLSYSTASDNNTQSVAKGSTK 260
QY 241 PANKSENLOSLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFPIYSK 300
Db 261 PANKSENLOSLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFPIYSK 320
QY 301 LSALLEETARWVPIISGTSTVSTNAKNEVVSLSGSSNPSSLTTSKELSSASDGYIFN 360
Db 321 LSALLEETARWVPIISGTSTVSTNAKNEVVSLSGSSNPSSLTTSKELSSASDGYIFN 380
QY 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSLPIINPGTSHKHEED 420
Db 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSLPIINPGTSHKHEED 440
QY 421 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSINGLDSLS 480
Db 441 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSINGLDSLS 500
QY 481 HEQDYPGNKEMKDLKIEEIKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPID 540
Db 501 HEQDYPGNKEMKDLKIEEIKTAGIMKQYGVKRESIVVKNKNAIIPHGDDHADPID 560
QY 541 EHPVGIGSHSNYELFKPEEGVAKKEGKVTGTELTNVNLLKNSFNQNFPLANGQ 600
Db 561 EHPVGIGSHSNYELFKPEEGVAKKEGKVTGTELTNVNLLKNSFNQNFPLANGQ 620
QY 601 KRVSFPPELEKKGILNMLVKLITPDGKVLKESYGVGEGVGNIANFELDPYLPQGT 660
Db 621 KRVSFPPELEKKGILNMLVKLITPDGKVLKESYGVGEGVGNIANFELDPYLPQGT 680
QY 661 FKTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDAL 720
Db 681 FKTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDAL 740
QY 721 VRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNQSTIYVE 780
Db 741 VRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNQSTIYVE 800
QY 781 VPILKENDQKPSILPOFKRNKAQENSLDKVEPEPKTSEKVEKLSNLSNSTL 840
Db 801 VPILKENDQKPSILPOFKRNKAQENSLDKVEPEPKTSEKVEKLSNLSNSTL 860
QY 841 EEPVTPDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Db 861 EEPVTPDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
QY 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVGSDPM 960
Db 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVGSDPM 980
QY 961 LDPALAEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
Db 981 LDPALAEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIKKNLSDLIA 1039

RESULT 3
D98004
Histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:Cross-references: GB:AE007317; PIDN:ANK99864.1; PID:g15458682; GSPDB:GNO0174
C:Genetics:
A:Gene: phpA
C:Superfamily: *Streptococcus agalactiae* hypothetical 92.4k protein

Query Match	24.3%	Score 1292.5;	DB 2;	Length 855;
Best Local Similarity	40.9%	Pred. No. 3.6e-56;		
Matches 304;	Conservative	99;	Mismatches 182;	Indels 159;
				Gaps 21;

Query Match 24.3%; Score 1292.5; DB 2; Length 855;
Best Local Similarity 40.9%; Pred. No. 3.6e-56;
Matches 304; Conservative 98; Mismatches 182; Indels 159; Gaps 21;

QY	1	CAYALNHRSOENKDNRRVSYVDGSSQSSKSENLTPOVQSKEGIAEQAEVIXITDGGYV	60
Db	31	CAYELGHLHQAOYKVENNRVSYIDGKQATQKTENLTPEVSKREGINAQVIXITDGGYV	90
QY	61	TSHGCHYHYNGKVPYDALFSELLMKDPNQLKDAIVNEVKGYIIVDGYKYYVLKD	120
Db	91	TSHGCHYHYNGKVPYDAIISEELLMKDPNQLKDEDIISIEIKGGYVIVDGYKYYVLKD	150
QY	121	AAHADNVRTKEINRQKEAVKONE---KVNSENVAVARSQGRYTTNDGVVFVPADIIED	176
Db	151	AAHADNVRTKEEINRQKEHSQHREGGTPRNDCAVALARSQGRYTTDGGYIENASDIIED	210
QY	177	TGNAYIVPHGCHYHYIPKSDLSASELAALAAKHLAKKNQPSQISYSYSTADSN-----	228
Db	211	TGDAYIVPHGDHYHYIPKNLSASELAALAAKAFULSGRNLNSRTYRRQNSDTSRTNWVP	270
QY	229	-----NQTSVAKGTSKPANKSENQLSLKELXDPSSAQRYSESDGLVEFPAKTIISR	280
Db	271	SVSNPGTINTNTSNNSTNTSQASOSNDISLLKQLYKLPLSRHVESDGLFIDPAQITSR	330
QY	281	TPNGVAIVPHGDHYHYIPYSKLSALEEARIARVPTSGTGVSTNAKPN-----	328
Db	331	TANGVAVPHGDHYHYIPYSQLSPLEEKARIIPLYRSNHNWVPSRPEQSPQSTPEPSP	390
QY	329	--EYVSSLSLSSNPSSLTTSKE--LSSASQGYIFN-----PKDIVETAT---AYI	373
Db	391	SPQAPAPQAPSPAPIDEKLVKEAVRVKGVGYFEENGVPYRTPAKDLSAETAGIDSKL	450
QY	374	VRHGDHFHY-----IPKSN-----QIGOPTLPNNSLATFSPSL-----	406
Db	451	AKQESLSHKLCAKTDLPSSDRPEYNKAYDLARIHODLONKGRQVDFEALDNLLERLK	510
QY	407	-----PI-----NPGTSHEKHE-----EDGYGPDAN	427
Db	511	DVSSDKVKLYDDILAFAPIRHPERGLGKPNQAITYTDEIQVAKLAGKYTTEDGYIDPR	570
QY	428	RIIAEDSGFVMSHGDNHHFFFKDLTEBOIKAAOKHLEEVKTSNGLSDLSLSSHEQDYPG	487
Db	571	DITSDEGDAYVTPMHTSHWIKKDSLEAPRAAAQYAKE-----KGLTPPSTDHQP--SG	624
QY	488	NAKEMKOLDK-----KTEEKIAGIMKQYGVKRESIVYVNEKKNALIIYPHGHHPADPIDEH	542
Db	625	NT-PAKGAELYNRVKAAKVPDLDMYPYNLQ---YTVVRNGSLTIIPHYDHYH-----	673
QY	543	KPVGIGHSHSNYELFKPEGVAKKEKNKYTGELTNVMNLL-----KNSTFNNQN	593
Db	674	-----NIKFEWF--DEGL--YEAPKGYSLIEDLATVYKYVHPNRPNPHSDNGFCNAS	721
QY	594	FTL---ANGOKRVSFSPPPLEK	613
Db	722	DHVQRNKNQGADTNOTKEKPNEEK	744

RESULT 4
C97985
hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
E:Hosts: J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.
E:R.: LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y: P.: Sun, P.M.; Winkler, M.E.
J:Y: Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD
A:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	23.8%	Score 1263.5;	DB 2;	Length 853;
Best Local Similarity	41.6%	Pred. No. 9.6e-55;		
Matches 304; Conservative	96;	Mismatches 182;	Indels 149;	Gaps 24;

Qy	1	CAYALNHR	S-QENKDNRRVSYDGSOSSOKSENLPDPQVSQKEGTQAEQIVIKITDQGY	59
Db	20	CSYELGRHAGQAKESNRVSYIDGDAQGAENLTPEYSKREGINAEQIVIKITDQGY	79	
Qy	60	VTSBGDHYHYNGKVPYDAIFSELLMKDPNYQLKDSIYNEIKGGVYIKVDGKYVYVK	119	
Db	80	VTSBGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIYNEIKGGVYIKVDGKYVYVK	139	
Qy	120	DAAHADNVRKDETNROKQEHVKD-NEKVNNSAVARSGQRYTTINDGYVFNPAIDITDG	179	
Db	140	DAAHADNIRKKEELKROKQERSHNENSRADNAVAAAQAGRYTTDGYIFNASDIETDG	199	
Qy	179	NAYIVRHGGHYHYTPKSDLSASELAANAHAHLACKNMOPQSOLSYSSSTAASDNTQ----	235	
Db	200	DAYIVPHGDHYHYTPKSDLSASELAANAQAIAWNGK--OGSRPSSSSSSHNAPQPRSENH	257	
Qy	236	GSTKPA-----NKSENLOSILKELYDPSAQRYSBESDGLVEDPDAKITSRTNGVAIPHGDH	292	
Db	258	NLTVPTYHQNGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRTANGVAVPHGDH	317	
Qy	293	YHFTPYSKLSALEBKIAKMPVISGTGSTVSTNAKPN-----EYVSSLSGSL	338	
Db	318	YHFTPYSQLSPELEKLARIIPLYRSNHWPDSRPEQSPQSTPEPSQPAPNPQAP	377	
Qy	339	SNPSSLTTSKE-LSSASDGIEN-----PKDIVEETAT--AYIVRHGDHFHY--	382	
Db	378	SNPIDELVNEVRKVGCVFEENGVPRIIPAKDISAETHAAGIDSKLAKQESLSHLGA	437	
Qy	383	-----IPKSN-----QIGQTPLPNNSLATPSPSL-----	406	
Db	438	KKTDLPSSDREFYNKAYDILLARIHQDLNKNKQOVDFEALDNLLERLKDVSSDKVLVDD	497	
Qy	407	-----PI-----NPGTSHKEHE-----EDGYGFANRIIADESGFVM	439	
Db	498	ILAFAPIRHFERLGKNQAQIYTTDDIOVAKLAGYTTEDGYIFOPRDTISDEGDAYVT	557	
Qy	440	SHGDHNYFFKKDLTEBQIKAAQKHLBEVKTSHNGLDLSSSHQDPYGNAKEMKDK--	497	
Db	558	PHMTHSWIKKDSLSEBRAAAQAYAKE-----KGLTPPSTDHQD-SGNT-EAKGAETV	610	
Qy	498	---KTEKIAIGIKQYGVKKRESIVVNEKNAIYIPHGDDHHADPIDHKPVGIGHSHSNY	554	
Db	611	NRVAAAKKVLDRMPYNLQ---VTVEYKNGSLIIPHYDYH-----NKF	652	
Qy	555	ELFKPESGVAKKGNKYVTGEELTNVNL-----KNSTFNQNFTL---ANGQR	602	
Db	653	EWf---DEGL---YEAPKGYSLUEDLATYKYVVEHPNRPKSDNGFGNASDVQRNKQAD	708	
Qy	603	VSFSPFPELEK	613	

```
Db 709 TNQTEKNEEK 719
RESULT 5
E98004
Hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUR>
A:Cross-references: GB:AE007317; PID:RAK99865.1; PID:g15458663; GSPDB:GN00174
C:Genetics:
A:Gene: phtA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein
Query Match 23.7%; Score 1261; DB 2: Length 828;
Best Local Similarity 41.4%; Pred. No. 1.2e-54;
Matches 300; Conservative 95; Mismatches 179; Indels 150; Gaps 19;
QY 1 CAYALNHRSQENKNNRVSYDGSQSSQKSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db 32 CSYELGLYQARTVKENNRVSIDGKATQKTENLTDPDEVSKREGINAEQIVIKITDQGV 91
QY 61 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGYYIKVDGKYYVYVKD 120
Db 92 TSHGDHYHYNGKVPYDALFSEELLKDPNYKLKDEDIVNEVKGYYIKVDGKYYVYVKD 151
QY 121 AAHADNVRTKEINRQKQEHVKDNE----KVSNNVAVARSQGRYTTNDGYVFNPAIDIED 176
Db 152 AAHADNVRTKEINRQKQEHVQREGTGRNDGAVALARSGRYTTDDGYIFNASDIIED 211
QY 177 TGNAYIVPHGHHYHYTPKSDLSASELAALAHLAGKNNMQPSQSYSTASDN----- 228
Db 212 TGDAYIVPHGHHYHYTPKNELSASELAALAEAFLSGRNLSNRTYRQNSDNTSRTNWVP 271
QY 229 -----NTQSAVGKSTSKPANKSENLSLLKELYDSPSAQRYSESGLVDFPAKIIISR 280
Db 272 SVSNPGTTNTNTSNNSNTSQASQSDNDISLLKQLYKLPLSQRHVESDGLVDFPAQITSR 331
QY 281 TPNGVAIPHGDHYHYTPYKSLSALEEKIARWVP----- 313
Db 332 TARGVAVPHGDHYHYTPYQSMSELEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSP 391
QY 314 -----TSGGTSTVSTNAK-----PNEVYSSLSGLSSNPS 342
Db 392 GPQAPNPKLIDSNSSLSVLQVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQKE 451
QY 343 S-----LTTSKELSSASQGYIFNPK-DIVETATAYIVRHG--DHFHYIPK-----SNQI 389
Db 452 SVSHLTAKKNVAPROQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDEST 511
QY 390 GQPTLPNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFANDRIIA 431
Db 512 NKEKLVDDLLAFIAPITHPRLGRKPSQIEYTEDEVRIALQADKYTTSQGYIFDEHDIIS 571
QY 432 EDESGFVMSHGDNHYFFKFKDLTEEQIKAQKHLEE-----VKTSHNGLDSLSS 480
Db 572 DEGDAYVTPHMGSHHWIGKDSLSQKQKVAQAQYTKKGIPLPPSPDADVKANPTG-DSAAA 630
QY 481 HEQDYPGNAMKEDLDKIEEKTAGINKQGVKRESIVYVKNENKNAIYPHGDHHDADP-- 538
Db 631 IYNRVKG-----EKRIPLVRLPYWV--EHTVEYKNGNLII-PHKDHYHNKIFA 675
```

```
QY 539 -IDEH---KPVGIG-----HSHSNVELFKPREGVAKKBEKNKYVTGEELTNVNNLLKNSTFN 590
Db 676 WFDDHTYKAPNGYITLEDLFATIKYVVEHPDRPHSNDG---WGNASPHVLGKKDHSDEP 731
QY 591 NONE 594
Db 732 KNKF 735
RESULT 6
C95136
Conserved domain protein Spil175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1-802 <KUR>
A:Cross-references: GB:AE005672; PID:AAK75284.1; PID:g14972655; GSPDB:GN00164; TIGR:
A:Experimental source: Strain TIGR4
C:Genetics:
A:Gene: Spil175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein
Query Match 23.6%; Score 1255; DB 2: Length 802;
Best Local Similarity 41.3%; Pred. No. 2.3e-54;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps 19;
QY 1 CAYALNHRSQENKNNRVSYDGSQSSQKSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db 6 CSYELGLYQARTVKENNRVSIDGKATQKTENLTDPDEVSKREGINAEQIVIKITDQGV 65
QY 61 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGYYIKVDGKYYVYVKD 120
Db 66 TSHGDHYHYNGKVPYDAIIEELLKDPNYKLKDEDIVNEVKGYYIKVDGKYYVYVKD 125
QY 121 AAHADNVRTKEINRQKQEHVKDNE----KVSNNVAVARSQGRYTTNDGYVFNPAIDIED 176
Db 126 AAHADNVRTKEINRQKQEHVQREGTGRNDGAVALARSGRYTTDDGYIFNASDIIED 185
QY 177 TGNAYIVPHGHHYHYTPKSDLSASELAALAHLAGKNNMQPSQSYSTASDN----- 228
Db 186 TGDAYIVPHGHHYHYTPKNELSASELAALAEAFLSGRNLSNRTYRQNSDNTSRTNWVP 245
QY 229 -----NTQSAVGKSTSKPANKSENLSLLKELYDSPSAQRYSESGLVDFPAKIIISR 280
Db 246 SVSNPGTTNTNTSNNSNTSQASQSDNDISLLKQLYKLPLSQRHVESDGLVDFPAQITSR 305
QY 281 TPNGVAIPHGDHYHYTPYKSLSALEEKIARWVP----- 313
Db 306 TARGVAVPHGDHYHYTPYQSMSELEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSP 365
QY 314 -----ISGTGSTVSTNAK-----PNEVYSSLSGLSSNPS 342
Db 366 GPQAPNPKLIDSNSSLSVLQVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQKE 425
QY 343 S-----LTTSKELSSASQGYIFNPK-DIVETATAYIVRHG--DHFHYIPK-----SNQI 389
Db 426 SVSHLTAKKNVAPROQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDEST 485
QY 390 GQPTLPNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFANDRIIA 431
Db 486 NKEKLVDDLLAFIAPITHPRLGRKPSQIEYTEDEVRIALQADKYTTSQGYIFDEHDIIS 545
QY 432 EDESGFVMSHGDNHYFFKFKDLTEEQIKAQKHLEE-----VKTSHNGLDSLSS 480
```



```
Db 546 DGDGAVTPHMGHSHWICKDLSOKEVAAQAAYTKESIGILPEPSPDADVKANPTG--DGA 604
QY 481 HQDYPGNNAKEMKLDKIEKIACIMKQGVKRESIVVNKEKNALIIYPHGDHHDADP-- 538
Db 605 IYNRVKG-----EKRIPLRPLPVV--EHTVEVRNGNLII--PHKDHVHNKIFA 649
QY 539 -IDH--KPVGIG----HSHSNVELFPERGVAKKEKGVYTGELTNVYNLLKNSTFN 590
Db 650 WDDHTYKAPNGYTLDFATIKYVVEHDERPHSNDG---WGNASEHLVKKDHSDEP 705
QY 591 NONF 594
Db 706 NKNF 709

RESULT 7
B951136
conserved domain protein Spil174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B951136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
C:Accession: B951136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Spil174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.2%; Score 1235.5; DB 2; Length 819;
Best Local Similarity 41.0%; Pred. No. 2-28-53;
Matches 297; Conservative 102; Mismatches 183; Indels 143; Gaps 24;

QY 1 CAYALNQHRS-OENKDNRRVYVDGSSQSSSENLTDPQVSOKEGIAEQIVIKITDQGY 59
Db 20 CSYELGRYQAGDKKESRVAYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYTIKVDGKYYVYLK 119
Db 80 VTSBGDHYHYNGKVPYDAIIEELMKDPNQLKADIVNEIKGYYIKVNGKYYVYLK 139
QY 120 DAAHADNVRTKDEINROKQEHVKD-NEKVNNAVARSQGRYTTNDGYVFNPAITDGT 178
Db 140 DAAHADNRTKEEKIKRQKQEH-SINHHGGSDNQAVARAQGRYTTDGGYIFNASDIETG 199
QY 179 NAYIYPHGHYHYIPKSDLSASELAALAAKHAHLAGKNMPSQLSYSTASDNNTQ----SVAK 235
Db 200 DAYIYPHGDHYHYIPKNSLASSELAALAAEAYWNGK--QGSRRPSSSSSYNANPAQRLSENH 257
QY 236 GSTSXP-----NSENLSQLLKELYDSPAQRYSSESDGLVDFPAKTIISRTPGVATPHGDH 292
Db 258 NUTVTPTTHQNGENISSLLRELYAKPLSERHVESDGLIFDPAQTTSRTARGVAVPHGNH 317
QY 293 YHFIYPSKLSALEEKIARMPISGTGSTVSTNAKEN-----EVSSSLGSLSSNPSSL 344
Db 318 YHFIYEQMSELEKRIARIPLRYRSNHWVPDSRPEEPSQPTPEPSPQAPSPNIDE 377
QY 345 TTSKE-LSSASGQYFN-----PKDVEETAT---ATVIRGDIHFY-----IP 384
Db 378 KLVKEAVRVKGDGYFEENGVSRYIPAKDLSAETAAGDSKLAKQESLSHKLGTKKTDLP 437
QY 385 KSN-----QIGOPTLPNNSLATPPSPSL----- 406
```

```
Db 438 SSOREFYNKAYDILLARTHQDLLDNKGRQVDFEALDNLRLERLKDVSDDKVKVLIEDILAFLA 497
QY 407 PI-----NPTCSHEKE-----EDGYPDANRIIIEDESGFYVMSHGDHN 445
Db 498 PIRPELSPKPNQAIIITDDEIQVAKLAGYTTEDGYIFDPRDITSDGDAYVTPHWTSHS 557
QY 446 HYFFKDLTEEQIKAAQKHLEEVKTSINGLSDLSSEHQDYPPGNKAKEMKOLDK-----KIE 500
Db 558 HWTKDLSLSAEARAAQAAYAKE-----KGLTPPSTHDQD--SGNT-EAKGAEATYNRVAAA 610
QY 501 EKTAGIMKQGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNVELFKPE 560
Db 611 KKVLDRMPYNLQ---YTVEVKNGSLIIPHYDHYH-----NIRFEW--D 650
QY 561 EGVAKEGKNKVTYGEELTNVNNL-----KNSTFNQNFLL-----AKQKRVSFSP 608
Db 651 EGL--YEAPKGYTLEDLATVKYVVEHPNRPDSDNGFNASDHVQRNKGQADTQOTEK 708
QY 509 PELEK 613
Db 709 PSEK 713

RESULT 8
G95115
conserved hypothetical protein SpI003 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
C:Accession: G95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SpI003
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.2%; Score 1233; DB 2; Length 839;
Best Local Similarity 42.8%; Pred. No. 3e-53;
Matches 295; Conservative 90; Mismatches 164; Indels 140; Gaps 25;

QY 1 CAYALNQHRS-OENKDNRRVYVDGSSQSSSENLTDPQVSOKEGIAEQIVIKITDQGY 59
Db 20 CSYELGRYQAGDKKESRVYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYTIKVDKYYVYLK 119
Db 80 VTSBGDHYHYNGKVPYDAIIEELMKDPNQLKADIVNEIKGYYIKVVDGKYYVYLK 139
QY 120 DAAHADNVRTKDEINROKQEHVKDNKVNNSN---VAVARSQGRYTTNDGYVFNPAITDIED 176
Db 140 DAAHADNRTKEEKIKRQKQEH-SINHHGGSDNQAVARAQGRYTTDGGYIFNASDIETD 198
QY 177 TGNAYIYPHGHYHYIPKSDLSASELAALAAKHAHLAGKNMPSQLSYSTASDNNTQ----SV 233
Db 199 TGDAYIYPHGDHYHYIPKNSLASSELAALAAEAYWNGK--QGSRRPSSSSSYNANPAQRLSE 256
QY 234 AKGSTSXP-----NSENLSQLLKELYDSPAQRYSSESDGLVDFPAKTIISRTPGVATPHG 290
Db 257 NENUTVTPTTHQNGENISSLLRELYAKPLSERHVESDGLIFDPAQTTSRTARGVAVPHG 316
QY 291 DHHFIYPSKLSALEEKIARMPISGTGSTVSTNAKEN-----EVSSSLGS 336
Db 317 NHHFIYEQMSELEKRIARIPLRYRSNHWVPDSRPEEPSQPTPEPSPQAPNPQP 376
```

```
QY 337 LSSNPSSLTSE-LSSASDGIYFN-----PKDIVEETAT---AYIVRHGDHFHY- 382
DB 377 APSNPIDEKLVAEVRKVGDDGYVEENGVSRYIPAKDLSAETAGDLSKLAKQESLSHL 436
QY 383 -----IPKSN-----QIGQPTLPN-----NSLA-----TPSPSL--- 406
DB 437 GAKKTDLPSSDREFYNKAYDILLARIHQDILLDNKGRQVDPEALDNLRLKDVPSKVLV 496
QY 407 -----PI-----NPQTSSEKHE-----EDGYCFANRIIADESGF 437
DB 497 DOILAPLADIRPERLGKNAQITVDDIQAQVAKLAGKYTTEDGYIFDPDRITISDEGAY 556
QY 438 VMSHGDHNYHFKDILTEBOIRAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDKD 497
DB 557 VTPMTHSHWIKDSLSEAEARAAQAYAKE-----KGLTPPSTHQD-SGNT-EAKGAE 609
QY 498 -----KIEKIAIGIMQYGVKRESIVVNKKNALIIYPHGDHHDHPIDHKKPVGIGHSHS 552
DB 610 IYNRVAAKVPDLDRPNYLNQ---YTVKNGSLIPIHYDHYH-----NI 651
QY 553 NYELFPPEEGVAKKCKGKYVTCGELTNV 581
DB 652 KPEWF--DEGL--YEAPKGYTLEDLATV 676

RESULT 9
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Speilsteberg, B.; Rodzinski, E.; Martin, S.; Weber-Heynenann, J.; Schnitzler, N.; Luet
Infect. Immun. 67, 871-878, 1999
A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachment
A:Reference number: 224091; MUID:99115568; PMID:9916102
A:Accession: T46758
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPES>
A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AADI3797.1; PID:g4249624
A:Experimental source: strain R268
A:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 16.9%; Score 899.5; DB 2; Length 822;
Best Local Similarity 27.0%; Pred. No. 8.6e-37;
Matches 262; Conservative 144; Mismatches 297; Indels 267; Gaps 31;

QY 2 AYALNQHRSQENKDNRRVSVYDGSQSSQKS--ENLTPQVSOKEGIAEQIVIKITDQGY 59
DB 22 SYVLGKHMHGLATKNQIAYIDDSKGKYPKNTKMTDQISAEIGISAEQIVVKITDQGY 81
QY 60 VTSBGDHYHYNGKVPYDALFSEELLKMDPNYOLKADIVNEVKGYIKYVDGKYVYVLK 119
DB 82 VTSBGDHYHYNGKVPYDAIIEELLMTDPNHYFKQSDVINEILDGYIKVNGNYVYVLK 141
QY 120 DAADADNVTKDEINROKQEHVKD-NEKVNVA-----VARSQGYTTNDGYV 167
DB 142 PGSKRNITTKQIAEQVAKGTEKAKEGLAQVAHLSKEEVAAVNEAKRQGYTTDDGYI 201
QY 168 FNPADITDGNAYIVPHGGHYHYIPKSDLSASELAARAHLA---GKNMPSQLSYST 224
DB 202 FSPDTIIDLDGDAYLVPHGNHYHYIPKDLSPSELAAQAQYAKSQGRGARDS--DYRPT 259
QY 225 ASDNNTQSVAKGSKSPAN-----KSNLQSLKEL 255
DB 260 PAPGRKKAIPDVTNPNGOCHQPDNGYHAPPRNDASQNKHQDEKFKGTFKELLQOL 319
QY 256 YDPSQAQRYSESDGLVDFPAKIIISRTPNQVAIPHGDHGHFIPIYSKLSALEEKIARWPIS 315
DB 320 HRLDLKYRVEEDGLIFEPTQVICKSNAFYVYPHGDHYHIIIPRSQSLPELEADRYLAG 379
QY 316 GTGSTVSTNAKNEVVSSLGSLSSNPSSLTTSKELS-----SASD 355
```

```
DB 380 QTDNDN-----GSDHKS-----DKRYHTFLGHRKAYKGLDGPYDTS 423
QY 356 GYINPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPN--NSLATPSPSLPIMPPTS 413
DB 424 AYVESKESIHSVDKSGVTAKHGDHFHYI-GFGELEQVELDEVANWVKAKQADDELVAALD 482
QY 414 HEKHEEDGYFD---ANRIIADESGFVMSHGDNHYEFKKDLTEBOIRAAQKHLEEVK 469
DB 483 QEQKREKPL-FDTKKVSRKVTGKGVGIMPCKGDKIFYARYQLDLTQIAFAQELMLKD 541
QY 470 TSHNGDLSLSHEQDYPGNAKEMKDLKIEEKIAGIMQYGVKRESIVVNKKNALIIYP 529
DB 542 KKHYRYDI-----VDTGTEPLAVDVSSLPNHAGNATYDTCSSFVI-P 583
QY 530 HGDHHDHP---IDEKPVGIGHSHSNYELKPE--EGVAKKSGKNKYVTGCEELTNVNL 584
DB 584 HIDHIVVVPYSLTRNQIATI-----KYVMQHPEVRPDVMSKPGHE-ESGVIPTNVTPLD 637
QY 585 KNSTFNQNETLANGQKRVSFSPPELEKELGINMLVKLITPDCKVLEKVGKVFGEVGV 644
DB 638 KRACMPNQI-----IHSAEVQKALAE--RFAAPDGYIFD----- 672
QY 645 NIANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGDTY 704
DB 673 -----PROVLAKETF-----VWKGDSFSIPRA----- 694
QY 705 LRVPQFAVPGKTDALVRVDEFHGNAYLENNYKVGKIKLPIPKLNOGTTRTAGNKIPVT 764
DB 695 -----DSSLRINKSDL- 707
QY 765 FMANAYLDNSTYIVVPILEKEN-----QTDKPSILPQKRNKAQENSKLDEKVEEPT 819
DB 708 -----SOAEQOQACELLAKKAGADATDTPKEEQQ--ADKSNEN-----QOPSE 750
QY 820 SEKYEKELSETGNSNSTLEEVPTVD---PVOEKVAKFAESYGMKLENVLFNMDGTI 875
DB 751 ASKEEKE-----SDDFIDSLPDYGLDRAILEDHINLAQXANIDPKYLIFQPEGV 800
QY 876 ELYLPSGEVI 885
DB 801 QFYKNGELV 810

RESULT 10
F97985
hypothetical protein phxE-truncation [imported] - Streptococcus pneumoniae (strain R6
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.W.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:gl15458517; GSPDB:GN00174
C:Genetics:
A:Gene: phxE-truncation

Query Match 6.1%; Score 324.5; DB 2; Length 182;
Best Local Similarity 41.0%; Pred. No. 2e-09;
Matches 73; Conservative 26; Mismatches 60; Indels 19; Gaps 5;

QY 344 LTTSKELSSADSGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPN-NSLATP 402
DB 3 VTPNGVSAVDGQGVFNPDIVRDTGDAYIVRHGDHYHYIPKSLNNPPSHSNTTEVGSS 62
```

```

Db 444 AVV-----TDKGET--EYQPESPDTVVSDKGEPEQVAPLPYKGN-----I 482
QY 740 GEIK--LPIPKL-NOGTRTAGNKIPYTEMANALYDNOQSYIVVEVPILEKENOTDKPSIL 796
Db 483 EQVKPEPVKTEKQGEKPT--EVPV-----KPTETPVNNEGTTGTSI-- 527
QY 797 PQKRN---KAQENSKLDEKVEBKPTSEK-----VEKEKJSETGNST 835
Db 528 -QEAENPVQAEESTTINSEKV-SPDTSKNTGEVSSNPSSDSTTSVGESNRPENHDSKNEN 585
QY 836 SNSPLREVPVDPVQEKVAFESYGNKLENLVFNMDGTIELYLPSEGVKKKNADFTGE 895
Db 586 SEKTVEVP-VNP-----NEGTV-----GTSNGETE 611
QY 896 APOGNGENKPSGKV---STGTIVENQPTENK--ADSLPEAPNEKPKVPKENS---TDN 946
Db 612 KPQVPAEETQTSKIANENTGEVSNKPSDSKPPVEESNQPEKNGTATKPENSGNTTSEN 671
QY 947 GMLNPE---GNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVFNMDGTIELR 1003
Db 672 GQTEPEPSNGSTEDVSTESNTSNSNGNEEIKQENE-----LDPOKKVEEPEKLTIELR 724
QY 1004 LPSGEVIKKNLSD 1016
Db 725 -----NVSD 728

RESULT 12
H71621
serine/threonine-specific protein kinase (EC 2.7.1.-) PFB0150c - malaria parasite (Pl
C:Species: Plasmodium falciparum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: H71621
R:Gardner, M.J.; Rettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: H71621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2485 <GAR>
A:Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AACT71820.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0150c
C:Superfamily: malaria parasite serine/threonine-specific protein kinase PFB0150c; pr
C:Keywords: phosphotransferase
F:2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.3%; Score 228; DB 1; Length 2485;
Best Local Similarity 18.1%; Pred. No. 0.0068;
Matches 208; Conservative 149; Mismatches 392; Indels 400; Gaps 47;

QY 3 YALNQHRSQENKONNRVSYVDGSGSSQSKSENLPDQVSQKEGIAEQIVIKITDQGVVTS 62
Db 564 YKRRKSNNNNNNNNIS--SSSSSSKKKHVI---INKK-----ISS 601
QY 63 HGDRHYHYNGKVPYDALFSELLM-----86
Db 602 YNTHYKERKDSFENFLFFKEKILPSKKDTCVNERQKDLFPEKSNIEHKCVSFENNTSDD 661
QY 87 -----KDPNYQLKADADIVNEVRGGYIIRKYDGYVYLKDAAHADNVRTKDEINRQK 137
Db 662 ISSHSSVNNKPEPTFALKNNIRIHPKRENNIIYTSGRKSNFHVQKEKTVLLKKKKEIN--- 718
QY 138 QEHVKDNEKYNVAVARSQGRYTTNDGYVFNADI-----IEDTGNAIVPVCGHY 189
Db 719 DKNTFFSCLINHITT-----YTLQNGVKNKLNWLGIRDSIYKIDKKNMLKECYNGNN 772
QY 190 HYIPKPSLSASELAHAKHLAGKNMPSQLSYSSSTASDNNNTQSVAKGSTSKPANKSENLO 249

```

Db 773 DSNKKKKKKLSPS-CDIINDNTP-----YESDEKNNNNKMSKIDFNVKRSNLY 827
QY 250 SLLKELYDS--PSAORYSDGLVDPDAKIIISRTNGVAIPHGDHYHPIXPY- 299
Db 826 NLSNRDSTVDHMKYSEYI-----NIQRTKIVELSKRIRNKLYSMDEIFKVS 881
QY 300 -----KLSALEKERTARVPI----- 314
Db 882 LREKYIDNISNMERVYTKNEMINEKISKDDIILYPCDKNKLNSMCPVIENNISRE 941
QY 315 SGTGSTVSTNAKPE-----VSSLSGLSNPSSSLTSKELSSAS 354
Db 942 NEKNSVILKKNENMFCVGRLCHGMKNMNQDNIYDQNIKKNEEIKHDEYISRE 1001
QY 355 DGYIFNPKDVEETATYIVRHGDHFHYIPKSNQIQGTPLPNSLATSPSLPINPGTSH 414
Db 1002 EKNKYSK-----CIRNFDDYK-----EQV-----LSY 1025
QY 415 EXHEEDGYGFDAIRIIDE-----SGFVMSHGDHNYHFFKDLTEEQIKAQK 463
Db 1026 HTLDEDKKNDNNLIDMNEAIIETVNGVINIILDRKDNNS-----RKDMKEMEKEMK 1082
QY 464 HLEP--VKTSHGLDLSLSEHQDYPGNAKEMK-DLDKXIEEKIAGIMKO----- 509
Db 1083 KMEKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMKEMK 1137
QY 510 YGVKRESIVNNEKNAIYPHGDHHAADPIDEHK-----PVGIGHSHSNY----- 554
Db 1138 VDMDKELEIVNEKK--LITPFYNES-----DVHKNMNSINNNCKDDIINNILKEYVDNS 1191
QY 555 -----ELFKPEEGYAKKE-----GNKYIT--GEELTNVYLLNKNSTFNNONFT- 595
Db 1192 CLAQKEENIFRPLNKKDKVKNRKNKNIKTIIHNEEMKRIYQTTINKVFIYNENR 1251
QY 596 -----LANGOKRVSPFPPELEKLGINMLVKLITPDGKRVLEKVSQVFGVGN--IANF 649
Db 1252 YENFLIN--HUTYNPK-----NDLFKL-----SYKVSNNIRNLYIANK 1289
QY 650 ELDQPY-----LPQO--TFKYTIASKDYPE--VSQDTFTVPTSLAYKMASQIFYPF 698
Db 1290 HINNVDYNNKLYNNIYTLKYQVANIDNDHICKGGGLDYNNNISKECKNR----- 1343
QY 699 HAGDTYLRNPOFAPKGTDALVRVDFDEHGNAYL-----ENNYKVEIKLPIPK 748
Db 1344 -KDKTYL--NKIFHYKKKKDARFFINDEIGSDNYDIKKKYSNDENNYKLE----- 1393
QY 749 LNOGTTTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQTDKPSILPOKRNKAQENS 808
Db 1394 -----KNNISKNDEDMI-----PTLNSEHGNFPSCQPNLLEK--ST 1430
QY 809 KLDEKVEEPTSEKVEKELSTGNSNSTLEEVPTVDPVOEKVAKFAESYGMKLENVL 868
Db 1431 YIDLNLDSNMDDETEEKYFNVEN-----DLFNTKRWKFNFSKGNLFNKK 1479
QY 869 F----NWDGTILYLPSEGEVVKKNADFTG-----EAPQNGENKPSKGVSTG 914
Db 1480 FTVNSNEDGVFF-----KNMNLPRKNSNLSKLESVKNSNNCSNKKGGDDNIG 1531
QY 915 TVENQPTEN 923
Db 1532 NMENMNTN 1540
RESULT 13
B98002
IGA-specific metalloendopeptidase (BC 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: B98002
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: B98002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1963 <RUP>
A:Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:g15458662; GSPDB:GN00174
C:Gene: iga
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 4.2%; Score 223; DB 2; Length 1963;
Best Local Similarity 22.8%; Pred. No. 0.0083;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;
QY 186 GGHYHYIPKSDLSASELAALAAKHAHLAGKNMOPSLYS---SSTASDNNNTQSVAKGSTSKFA 242
Db 47 GVHYKYVADSELSEEE-----KKQLVYDIPTYVENDDETYLV----- 84
QY 243 NKSENLOSLLKELYDSPAORYSESOGLVDPDAKIIISRTNGVAIPHGDHYHPIXPYKLS 302
Db 85 -YKLNQNLAEEL--PNTGSKNERQALVAG---ASLAALGILI-----FAVSKKK 128
QY 303 ALEEKIARMPVPIGSGS--TVSTNAKPNVSVSSLSGSSNPSSSLTSKELSSASDGYIFN 360
Db 129 VKNKTVLHLVLVAGMGNGLVSVHALENHLL---LNYN-----TDYELTSGEK--LPL 176
QY 361 PKDVEETATAYIVRHGDHFHYIPKSNQIQGTPLPNSLATSPSLPINPGTSHKEHEED 420
Db 177 PKEISGYTYIGYI--KEGKTTSDFEVSNQ-----EKSAATPT-----KQKV 216
QY 421 GYGFDAIRIIDEDESQVMSHGDHNYHFFKDLTEEQIKAQKHLDEVKTSNGLDLSLS 480
Db 217 DYNVTPNV---DRPSIVQAIQGTTPVSSSTKP-TEVOVVEKPTSTELINPRKEEKSSDS 272
QY 481 HEQDYPGNAXEMKDLKKIEKTAGIMKQYGVKRESIVNNEKNAIYPHGDHHAADPID 540
Db 273 QEQ-----LAHKNLTKKEEKIS-----PKERTGV-----NTLNQD 305
QY 541 E-----HKPVGIGHSHSNYELPKPEGVAKKEGKGVYTGTEELTNVYLLNKNSTENNPF 594
Db 306 EVLSGQINKP-----ELLYRETIE---TKIDFOEEI-----QENP 338
QY 595 TLANGQRVSPFSFPELEKLGINM-LVKLIT-----PDGKVLKVSQ 636
Db 339 DLAEQTVRV-----KQEGKLGKVKELVIRIPSVNKEEVSREIVSTSTAPSPRIVEKGT 392
QY 637 K--VFGE-----GV-----GNTANFELDQPIPLQGTFTKYTIASKDYPEVSYDGTFTVPT 683
Db 393 KTVQIKQEPETGVEHKDVQSGAIVEPAI-QPELP-----EAVSDKQGEVQV---TLPE 443
QY 684 SLAYKMASQITIFYPFHAGDTYLRVNQO---FAVPKGTDALVRVDFDEHGNAYLNNYKV 739
Db 444 AVV-----TDKGET--EQVESPDTVYSDKGEPEQVAPLPEYKGN-----I 482
QY 740 GEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQTDKPSIL 796
Db 483 EQVKPETPVKTKDQGPKEK--EEVPV-----KPTETPVNPNREGTGTISI- 527
QY 797 POFKRN---RAQENSKLDEKVEEPTSEKVS---EKKKLSSETGNS 836
Db 528 -QEAENPVQAEESTTNSSEKVSPTSSSENTGEVSSNPSSDSTTSVGENSKPEHDSKNENS 586
QY 837 NSTLEEVPTVDPVOEKVAKFAESYGMKLENVLNMDQTIIELYLPSEGEVVKKNADFTGEA 896
Db 587 EKTVEEVP-VNP-----NEGTVE-----GTSNQETEX 612
QY 897 PQNGENKPSKGVK---STGTVENQPTENKP--ADSLPEAPNEKPKVPKPNK-----TDNG 947
Db 613 PVQPAETQTNISGRANIENTGEVSKNFSKDPKPPVEESNQPEKNGKATKPNSENSGNTSNG 672

QY 948 MLNPE 952
DB 673 QTEPE 677

RESULT 14
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodium
A:Reference number: A54498; MUID:88142999; PMID:2449612
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PEP>
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 4.1%; Score 220; DB 2; Length 1701;
Best Local Similarity 20.3%; Pred. No. 0.0094;
Matches 246; Conservative 171; Mismatches 395; Indels 402; Gaps 65;

QY 3 YALNHRSEN-----KDNRRSVYDGSQSSQKSENITPPQVSKQKGIQAEQIVIKITDQG 58
DB 519 YTYNVEKQRYNNKFKSSNNSVYVNV---QKLKALKSYLED-YSLRKG-----561
QY 59 YVTSKGDHYHYNGKVPYDA---LFESEL-----LMKDPNYOLKDAIVN- 100
DB 562 ---SEKDFNHYTLKTGLADIKLTKTEEKSSKILENFKGLTHSANASLEVSDIVKL 618
QY 101 EVKGGVII-KVGG--KYIYVLKDAHADNVRT-----KDEINRQKQ- 138
DB 619 QVQVLLIKKIEDLRKIELFLKNAOLKDSIHVPNIYKPNQKPEPYLIYLVKKEVDKLEF 678
QY 139 -----EHWKDNKENVNVA---VARSGRYTNDG-----YVFNPAIIIEDT 177
DB 679 IPKVKDLKKEQAVLSITQPLVASE---TTEDGGHSTHTLSQGETEVTETET 735
QY 178 GNAYIVPHGGHYHI-----PKSDLSASLAAKAHLAKNNQPSQLSYSTASDNNTQS 232
DB 736 -----VGHTTTVTITLPPKEESAPKVKVYVENSIEHK-----SNDNSQA 774
QY 233 VAKGSTSKPANK-----SENLSQLLKELYD-SPSAQRYSSEGLVDFPAK 276
DB 775 LRTVYLLKLDEFLTKSYCHYIIVNSNSMDQKULEVNLTPPEENELKS-----CDPLD 830
QY 277 IISRTPNGVAIPHG-----DHYHFTPYSK-----LSALEEK-----IARWVPISG 316
DB 831 LFNIGNNIPAMVSLYDSMNIDQLHLFFELYOKEMYIYLLHKEENHIKKLEEQKQITG 890
QY 317 TGTSTVST-NAKNEVSVSLGSLSSPSSLTTSKELSSASDGVIFNPKOIVETATAYIVR 375
DB 891 TSTSTSPGNTVNTAQSATHSNQSQNSASS---TNTQNG-----VAVSSGPAVVEE 940
QY 376 HGDHFHYIPKSNQI-GOPTLPN-NSLAPPSPSLPINPOTSHKEHEEDGYGDFDANRIAE 432
DB 941 SHDPLTVLSINDLKGIVSLNLGNKTKVPNP-LIIST-TEKEKYEN-----ILKN 990
QY 433 DSGGFVMSHGDNNHVFKKD-----LTEEIKAAQKHEEYKTS-----HN-----GL 475
DB 991 NDTYF-----NDDIKQFVKSNSKVIITGLTQKNALNDEIKLKDITLQSFGLYNNYKLL 1046
QY 476 DSISSHEQDYPGNAREMD----LDKKIEKIAKIMKQCVKRE-SIVYNKEKNAII-----527
DB 1047 DRUFNKKELGODKMOIKLTKLLEKLSKLSLNNPHNVLFNFSVFFNKKKEAEIAETE 1106
QY 528 -----YPHGDHHDHAPIDRHKPVGI---GSHSHSNYELFKPEEGVAKKE 567

DB 1107 NTLENTKILLAHYKGLVKYNGE-----SSPLKTLSEVSIQTEDNYANLEKFRALSIDKGL 1163
QY 568 GNKVVYTGEE-----LTNVNLLKNSTFNQNTLANGOKRYSEFSPPELEKLGIGNM 619
DB 1164 NDNHLGKKKLSFLSSGLHLHITELK-EVINKNNT-----GNSPSENKKA--VNE 1211
QY 620 LVKL---ITPDGKYLEKSGVKFGEVGNIANFELDQPYLPGQTFKYTI-----ASKDYP 671
DB 1212 ALKSYENFLPEAKVTVTVP-----PQPDVTPSPLSYRVSSGSGSTREET 1256
QY 672 EVSYDGTITVPTSLAYKKNASQTIFFYPFHAGDTYLRVNPQFAVPKGTG-----ALV 721
DB 1257 QIPTSG-----SLTELOQVVQNYDEDDSLVLPIFGESDNDYLDQVVTGEAIS 1310
QY 722 RVFDEHGNAYLENNYKVGKIKPLKLNQGTTRTAGKNIP---VTFMANA-----769
DB 1311 VTMDNILSG--FENEYDVILK-PL---AGYRSLKKQIEKNITFNLDILNSLK 1363
QY 770 ---YLD-----NQSTIVEVPILEKENQTDKPSILPOFRNK-----A 804
DB 1364 KRKYELDVLESMLQFKHISSEYIIIE-DSFKLLNSEQNTLLKSYKIKESVENDIKFA 1422
QY 805 QENSKLDEKV-----EPEKTSKVEKEKLSGTNSTLEEVPTVDPQVEKVAKE 856
DB 1423 QEGISYIEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTPS-----PAKDEQKKSKF 1477
QY 857 A-----ESYGMKLENVLFNM-----DGTIELYL 879
DB 1478 LPFLTNIETLYNNLVNKIDDDYILNLKARINDCNVEKDEAHVKITKSLDKAIDKIDLF- 1536
QY 880 PSGEVYKKNMADFTGEAPQNGENKPSNGK-VSTGTGVENOPT-----ENKPADSIPE 931
DB 1537 -----KNTNDFPAKKLINDTKKMDLGLSTGLVONFPNTIISKLEEGFQDMINT 1589
QY 932 APNE--KPVKPSNS 943
DB 1590 SOHQCVKKQCPNS 1603

RESULT 15
T30944
surface protein precursor - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30944
R:Shankar, V.; Baghdavan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A:Title: Infection-Derived Enterococcus faecalis strains are enriched in esp, a gene
A:Reference number: Z20943; MUID:99081742; PMID:9864215
A:Accession: T30944
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1873 <SHA>
A:Cross-references: EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AAD09858.1

Query Match 4.1%; Score 218; DB 2; Length 1873;
Best Local Similarity 20.8%; Pred. No. 0.014;
Matches 215; Conservative 122; Mismatches 377; Indels 322; Gaps 49;

QY 71 NGKYPYDALFSEELLKMPNQLKDADIVNEVKGYIIVKDGKYYVYLKDAHAD-----N 126
DB 593 KNNUPEDAEYSWKTEPDTNSVNTDSTKGIVTVYKIGNRTFDVDVEFAVRASQAMENDATYVP 652
QY 127 VRTKDEINRQKQEHVKQNEKYNVNAVARSQRYTNDGY---VFNPADIIETGNAYI 182
DB 653 ITTTPETTIQSGKPTFDKPDVP-----LANDAFSVLDVYN-----KDFGNASV 695
QY 183 VPHGHYHYIFPKSDLSASELAAAKHAHLAKNNQPSQLSYSTASDNNTQSVAKGSTSKPA 242
DB 696 DANTGIVTFTPAKGVGESEPTGTI-----PIKIVY-----QDGSVGTDLAV 738
QY 243 NKSENLSQLLKELYDPSAQ-----RYSESGLVDFPAKIISRTPNGVAIPHGDHYH 294

Db 739 TVSKN-----IYENPGENIPAGYHKVFTTAGEGTSIESGTTFAVKDGYSLPE----- 786
QY 295 FIPYKLSALEEKIARMPISGTGSTVSTNAK-PNEVSSLSLSSNPSSLTTSKELSSA 353
Db 787 ----DKLPVLKAKDG-----YTDKWPKE-----ATQPIKADDTFVSSA 822
QY 354 S--DGYIFNPKD-----IVEETATAYIVRHGDHFIYIPKSNQIGQPTL 394
Db 823 TKLDDIENPGONIPAGYHKVFTTAGEGTSIESGTTFAVKDG-----VSL 868
QY 395 PNKSLATSPSPINPGTSHEKHEEDGYGFDANRIIAEDESQFVMS-----HGDH 444
Db 869 PEDKL-----PVLKADGYTDKWPKE-----EATQPIKADDTFVSSATKLDIENPGDN 919
QY 445 -----NHYFFKPD-----LTFEQ-----KAAQKHLE----- 466
Db 920 IPAGYHKVFTTAGEGTSIESGTTFAVKDGYSLPEDEKLVKAKDGYTDKWPGEATOPI 979
QY 467 ----EVTSHNGLDLSLSSHEODYPCNAKEMKDLOKIEEKIAGIMKOYGVKRESIVWNK 521
Db 980 KADDTFVSSAHLDDKSDADKXTPGQKVTTELNKE-PDASEGINKKDLKPKDAKYTWK 1038
QY 522 E-----KNAIYPHGDHHDPIDEHKPVGIGHSHSNYELFKPE-EGVAKKEGK 570
Db 1039 EKVDISTAGNKKGTVVVYSD-GSSDEVE--VDVIVTDNRSDADKYEPTVEGEVEIGK 1095
QY 571 VYTGEBELTNVNLKKNSTFNN-----QNFTLANGOKRVSFSPPELEKKGILNMLVKL 623
Db 1096 VDLTDNVTNLPTLPQGTIVTDVTPGGTIDTNPNGVEGVIETYPD-----GTRDTVKV 1149
QY 624 ---ITPDGKVLKSVGKVFEGV-----GNIANFELDQYLPQGTIFYTIASKDYP 671
Db 1150 PVEVTDNRSDADKYEPTVEGEKVEIGKVDLTDNVTNL-----PTLPQGT---TVT----- 1197
QY 672 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFDEHGNA 731
Db 1198 DVTPGGTIDTNPNGVEGVIETPD-----TY-----PDGKOTVKVPVEVTDNR 1239
QY 732 YLENNY-----KV---GEIKLP-----IPKLNQGT---RTAGNKIPVTFMANAYLDN 773
Db 1240 SDADKYEPTVEGEKVEIGKVDLTDNVTNLPTLPQGTIVTDVTPGGTIDTNPNGVEGI 1299
QY 774 QSTY-----IVEVPILEKENQTDKPSILPOKRNKAQENSKLDEKVEEPTSEKVEKEK 827
Db 1300 EYTPDGTKDVKVPVEVTDNRSDADKXTPMVEGEKVEIGKVD----- 1343
QY 828 LSETGNSTSLTEEVPDPOEKVAKFAESYGMKLENVFNMDGTIELYLPFG--EVI 885
Db 1344 --LTDNVTNLPTLPQGTIVTDVTPG-----GTIDTNPNGVEGVIETVYPDGTKDVT 1393
QY 886 K-----KNWADTGEAPQGNKPNSENGKVS-TGTVENQPTENKPADSLPEAPNEKPV 938
Db 1394 KVPVEVTDNRSDADKYEPTVEGE-KVEIGKVDLTDNVTNLPT-----LPQGTIVTDV 1445
QY 939 KPNSTDNGLNPEGN 954
Db 1446 TPGGTIDT---NTPGN 1458

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 ; Search time 11.9531 Seconds
(without alignments)

3535.857 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315

Sequence: 1 CAYALNQHRSGENKNNRVS.....IELRLPSGEVIKKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	221	4.2	1701	1 MSP1_PLAFM	P08569 plasmodium
2	220	4.1	1701	1 MSP1_PLAFM	P13819 plasmodium
3	205.5	3.9	1183	1 CRA_STAUA	Q53634 staphylococ
4	203.5	3.8	1849	1 IGA4_HAEIN	P45386 haemophilus
5	199	3.7	1726	1 MSP1_PLAFM	P04934 plasmodium
6	199	3.7	1726	1 MSP1_PLAFM	P50495 plasmodium
7	195.5	3.7	1637	1 MSP1_STAUA	P80544 staphylococ
8	189.5	3.6	1531	1 Y038_CAEEL	Q09459 caenorhabdi
9	186.5	3.5	1570	1 P3K1_DICDI	P54673 dictyosteli
10	185	3.5	2195	1 SC16_YEAST	P48415 saccharomyc
11	184.5	3.5	1466	1 SPA2_YEAST	P23201 saccharomyc
12	184	3.5	3381	1 PGCV_BOVIN	P81282 bos taurus
13	181.5	3.4	1682	1 MSP1_PLAF3	P19598 plasmodium
14	179.5	3.4	1636	1 BUD3_YEAST	P25558 saccharomyc
15	178.5	3.4	2738	1 PGCV_RAT	Q08764 rattus norv
16	177	3.3	1702	1 IGA2_HAEIN	P45384 haemophilus
17	176.5	3.3	1139	1 HMI1_MYCGE	Q49413 mycoplasma
18	176.5	3.3	1658	1 YMF7_YEAST	Q03661 saccharomyc
19	175.5	3.3	2748	1 NMU1_YEAST	Q00402 saccharomyc
20	174.5	3.3	1577	1 HLYA_PROMI	P16466 proteus mir
21	174.5	3.3	1807	1 VTA2_XENLA	P18709 xenopus lae
22	174.5	3.3	2485	1 PTND_HUMAN	Q12923 homo sapien
23	173	3.3	1694	1 IGA0_HAEIN	P4969 haemophilus
24	173	3.3	1790	1 USO1_YEAST	P25386 saccharomyc
25	172.5	3.2	1142	1 GIN4_YEAST	Q12263 saccharomyc
26	172.5	3.2	1532	1 IGA_NEIGO	P09790 neisseria g
27	172.5	3.2	1781	1 AKAC_HUMAN	Q02952 homo sapien
28	172	3.2	817	1 YG4A_YEAST	P46949 saccharomyc
29	172	3.2	2869	1 RBP1_PLAFB	Q00798 plasmodium
30	171	3.2	710	1 L178_ARAPH	Q06738 arabidopsis
31	169.5	3.2	1167	1 SCAL_STRPY	P15926 streptococ
32	168.5	3.2	818	1 PTK2_YEAST	P47116 saccharomyc
33	168.5	3.2	1435	1 EBAL_PLAFM	P19214 plasmodium

ALIGNMENTS

RESULT 1

ID	MSP1_PLAFM	STANDARD	PRT	1701 AA
AC	P08569;			
DT	01-AUG-1998 (Rel. 08, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)			
DE	(PMWSA) (P190).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI-Taxid=70153;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88011243; PubMed=3079521;			
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;			
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";			
RT	J. Mol. Biol. 195:273-287(1987).			
RN	[2]			
RP	REVISIONS TO 1403; 1569 AND 1629.			
RA	Tanabe K.;			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-115 FROM N.A.			
RX	MEDLINE=86136024; PubMed=3004972;			
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,			
RA	Stunnenberg H., Bujard H.;			
RT	"Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";			
RT	EMBO J. 4:3823-3829(1985).			
RL	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).			
CC	-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDa, 42 KDa AND 19 KDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X05624; CAA29112.1;			
DR	PIR; A26868; A26868.			
DR	PIR; B25120; B25120.			
DR	InterPro; IPR000561; EGF-like.			
DR	Pfam; PF00008; EGF; 1.			
DR	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
KW	SIGNAL 1 19 POTENTIAL.			
FT	CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.			
FT	CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).			

P07866 saccharomyc
P04932 plasmodium
Q01484 homo sapien
P04933 plasmodium
Q01550 xenopus lae
P52081 staphylococ
P27951 streptococ
P47035 saccharomyc
Q06105 saccharomyc
Q04958 saccharomyc
P45385 haemophilus
Q09904 schizosacch

1 ITE1_YEAST
1 MSP1_PLAFK
1 ANK2_HUMAN
1 MSP1_PLAFM
1 MSP1_PLAFM
1 TANA_XENLA
1 ATIL_STAUA
1 BAG_STRAG
1 YJH6_YEAST
1 S160_YEAST
1 YMF9_YEAST
1 YMF9_YEAST
1 IGA3_HAEIN
1 NI24_SCHPO

34 168.5 3.2 1435
35 168 3.2 1630
36 168 3.2 3924
37 167 3.1 1639
38 166 3.1 1744
39 165.5 3.1 1256
40 164.5 3.1 1164
41 164.5 3.1 1189
42 164 3.1 1222
43 164 3.1 1679
44 163.5 3.1 1545
45 163 3.1 1159

Query Match 4.2%; Score 221; DB 1; Length 1701;
 Best Local Similarity 20.3%; Pred. No. 0.0085;
 Matches 246; Conservative 172; Mismatches 394; Indels 402; Gaps 65;

FT CARBOHYD 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

QY 3 YALNQRHSEEN-----KDNRRSVSDGSSQSSQSENLTPOVQSKQEGIQAEQIVIKITDOG 58
 DB 519 YTYNVEKQRYNNKFNSSNNVNV---QKLKALSYLED-YSLRKG1-----561
 QY 59 YVTSHGDDHYHNGKVPYDA---LFSEEL-----LMKDPNYQLKDADIVN- 100
 DB 562 ---SEKDFNHYHLYTLTGLEADIKLKEETLEKSSSENKLEKNFKGLTHSASNASLEVSIVKL 618
 QY 101 EVKGGVII-KVDG-KYYVYLKDAHADNVRT-----KDEINRQKO- 138
 DB 619 QVQKVLIIKKIEDLRKLELFLKNAQLKDSIHVPNIYKQPKPEYYLVLKKEVDKLEF 678
 QY 139 -----EHVKDNEKVNVA---VARSGRYTTNDG-----YVENPADIIEDT 177
 DB 679 IPVKOMLKEQAVLSITQPLVAASE---TTEDGGHSTHTLSQSETVETETVETET 735
 QY 178 GNAYIVPHGGHYHYI-----PKSLSASLAAKAAHLAKGNOPSOVSSTASDNNTOS 232
 DB 736 -----VCHTQTTVTITLPPKESAPKVKVYVENSIEHK-----SNDNSQA 774
 QY 233 VAKGSTSKPANK-----SENLOSLLKELYD-SPSAORYSESDGLVFDPAK 276
 DB 775 LTTVTYLLKLEDFLTKSYCHKYLVSNSMDOKLEVLNLIPEERELKS-----CDPLD 830
 QY 277 IISRTPNQVAIPHG-----DHYHFTIPYSK-----LSALEEK-----IARMVPISG 316
 DB 831 LLFNQNNIPAMSLYDSNMNDLQHLFFELYQKEMYYLHLKKEENHKKLLEEQKQITG 890
 QY 317 TGSTVST-NAKPNEVSSLSGSSNSPSSLTTSKELSSADGVIENPKDIVEETATAYIVR 375
 DB 891 TSSTSPGNTVNTAQSATHSNQSOOSNASS---TNTONG-----VAVSSGPAVVEE 940
 QY 376 HGHDFHYIPKSNQI-GOPTLPN-NSLAPSPSLPINPGTSHKHEEDGYGDFDANRIAE 432
 DB 941 SHDPLTVLSTISNDLKGIVSLNLGNKTKYPNP-LTIST-TEMEKFEYEN-----ILKN 990
 QY 433 DESGFVMSGHNNHYFFKDK-----LTBEQIKAAQKHLEEVKTS-----HN-----GL 475
 DB 991 NDTYF-----NDDIKQFVKSNSKVIITGLTQKNALNDEIKKLADTQLQSFYLYNKYKLK 1046
 QY 476 DLSLSSHEDYPYNAKEMKD---LDKKIEEKIAGIMQYQVKRE-SIWNKERNKAI1-----527
 DB 1047 DRFPNKKELGQDKMQLKTLTLLKEQLESLNLSNPHNVQNFVFFNKKKEAEIATEE 1106
 QY 528 -----YPHGDHHRADPIDEHKPVG1---GHSHSNYELFKPEEGVAKKE 567
 DB 1107 NTLENTKILLKHYKGLVKKYNGE---SSPLKTLSEVSIQTDENYANLEKFRALSKIDGKL 1163
 QY 568 GNKYVTGEE-----LTWVNLKNTSTNNONFTLANQKRVSPSFPELEKGLKGINM 619
 DB 1164 NDNLHLGKKKLSFLSSGLHLHILTELK-EVILKNKYT-----GNSPSENKK--VNE 1211
 QY 620 LVKL---ITPDGKYLEKSVKGFEGVGNFTANFELDPQYLPQGTFTYTI-----ASKDVP 671
 DB 1212 ALKSYENFFPEAKYTTVTVP-----POPDVTPSPLSVRVSGSGSKKEET 1256

QY 672 EVSYDGTFTVPTSIAYKVASQTIYPFPHAGDTYLRVNPQFAVPKGT-----ALV 721
 DB 1257 QIPTSG-----SLTLEQQVQVOLQNYDEEDDSLVLPFIEGSEDNDEYLDQVVTGEAIS 1310
 QY 722 RVFDEFHGNAYLENNYKVGELKPIPKLNOGTFTAGNKIP---VTFMANA-----769
 DB 1311 VTMDNILSG--FENEYDVIVLK-PL-----AGVYRSKAKQTEKNITITFNLNLDILASRLK 1363
 QY 770 ---YLD-----NQSTIVVEVPILEKENQDQKPSILPQPKRKNK-----A 804
 DB 1364 KRKYFLDVLDSLMQFKHISSENYIIE-DSFKLLNSQKNTLLSKYKIKESVENDIKFA 1422
 QY 805 QENSKLEKV-----EETSEKVEKEKLSSETGNTSTNSLEEVPTVDPQVKVAKF 856
 DB 1423 QEGISYVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPS-----PAKTDQKESKF 1477
 QY 857 A-----ESYGMKLENVLFNN-----DGTIELYL 879
 DB 1478 LPFLTNTIELNVLNKKDDYLLINKAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF- 1536
 QY 880 PSGEVYKKNMADFTGEAPQNGENKPSSENGK-VSTGTVENQPT-----ENKPADSLPE 931
 DB 1537 -----KNTNDFEAIKRLINDDTKMDLGLKLLSTGLVQNFPTIISKLEIEGKFDQDMLNI 1589
 QY 932 APNE--KPVKPPENS 943
 DB 1590 SHOCVKKQCPENS 1603

RESULT 2
 MSPL_PLAFAF STANDARD; PRT; 1701 AA.
 AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMW5A).
 DE MSP-1
 GN MSP-1
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88142999; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G., Brown G.V., Anders R.F., Kemp D.J.;
 RA "Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum";
 RT Mol. Biochem. Parasitol. 27:291-302(1988).
 CC CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).
 CC CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC CC EMBL; M19143; AAA29653.1;
 DR PIR; A54498; A54498.
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.


```

FT CARBOHYD 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 3920B75E73D38552 CRC64;

Query Match 4.1%; Score 220; DB 1; Length 1701;
Best Local Similarity 20.3%; Pred. No. 0.0094;
Matches 246; Conservative 171; Mismatches 395; Indels 402; Gaps 65;

QY 3 YALNHRQSEN---KDNRRSYVDSQSQSKSENLTPOVSKQEGIOAEIVIKITDQG 58
Db 519 YTYNVEKQRYNKNFSSNNSVZNV---QKLKALSYLED-YSLRKG1-----561

QY 59 YVTSHGCHYHYNGKVPYDA---LFSEEL-----LMKDPNYQLKDADIVN- 100
Db 562 ---SEKDFNHYTLTKGLEADIKKITEEKSSKENILEKNFKGLTHSANALESVDIVKL 618

QY 101 EVKGGVII-KVDS--KYVYVLKDAHADNVRT-----KDEINRQKO- 138
Db 619 QVQKVLKIKIEDLRKIELFLKNAQKDSIHVPNIYKPNKPEPYLIIVLKKEVDKLEF 678

QY 139 -----EHWKDNKVNVA---VARSQGYTTNDG-----YVFNPAIDET 177
Db 679 IPKVKMLKKEQAVLSSITQPLVAASE---TTEDGCHSTLSQSGETEVTETEVTEET 735

QY 178 GNAYIVPHGGHYHI-----PKSDLSASBLAAKAHLAGKNQWPSQLSYSSFASDNNTQS 232
Db 736 -----VGHITTVITLPPKEESAPKVKVVSIEHK-----SDNNSOA 774

QY 233 VAKGTSKPANK-----SENLSLLKELYD-SPSAQRYSESGLVDFPAK 276
Db 775 LKTVYVKKLDEFLAKSYCHKRYVILSVNSMDQKLEVINLTPPEENELKS-----CDPLD 830

QY 277 IISRTNGVAIPHG-----DWHETPYSK-----LSALEEK-----IARVVPISG 316
Db 831 LFNIONTPAMSYLSDSNIDLOHFLFELYOKEMYIYLHKKEENHKKLEEQKQITG 890

QY 317 TOSTVST-NAKPNEVYSSGLSSNPSSLTTSKELSSASDGYIFNPKOIVETATAYIVR 375
Db 891 TSSTSPGNTVNTAQSAHNSQOQSNASS---TNTQNG-----VAVSSGPAVVEE 940

QY 376 HGDHFHYIPKSNQI-GQPTLPN--NSLATPSPSLINPGTSHKEHEDGYGDANRIATAE 432
Db 941 SHDPLTVLSINDLKGIVSLNLGNKTKVPNP-LTIST-TEMEKPYEN-----ILKN 990

QY 433 DESGFVMSHGDHNVHFFKKD-----LTEQIKAAQKHEEVKTS-----HN-----GL 475
Db 991 NDTYF-----NDDIQFVKNSKSVITGLIETQKNALNDEIKLKDLQLSFDLYNKYKLL 1046

QY 476 DLSLSSHEQDYPGNAMKMDK---LDKKIEPKIAGIMKQGVKRE-SIVNKEKNATII-----527
Db 1047 DRLFNKKKELGQDKWQIKKLTLLKQLESKLSLNNPHNVLFNSVFFPNKKKEABIAETE 1106

QY 528 -----YPHGDHHDADIDHFKPVG1-GHSHSNYELFKPEEGVAKKE 567
Db 1107 NTLENTKILLKHYKGLVKYNYNGE---SSPLKTLSEVSTQEDNYANLEKFRALSXIDOKL 1163

QY 568 GKNVYTGEE-----LTNVNVLKNSTFNQNFETLANGQKRVSFSPPELEKKLGINM 619
Db 1164 NDNLHGLKKKLSFSLGSLHLLTELK-EVINKKNVT-----GNSPSENKKK--VNE 1211

QY 620 LVKL---ITPDCKVLEKVSQKVFGEVGNIANFELDQYLPQCTKYII-----ASKDYP 671

```

```

Db 1212 ALKSVENELPEAKVTIVVTP-----POPDVTPSLSVRSVSGSSGKTEET 1256
QY 672 EYSDGTGTTVPTSLAYKMASQTIFYPFHAGTYLRVNPQFAVPKGTD-----ALV 721
Db 1257 QIPTSG-----SLLTEQQVVLQNYDEEDDSLVLP1FGESEDNDEYLDQVVTGEAIS 1310
QY 722 RVDEFHGNAYLNNYKVEIKLPIKLNQCTTTAGNKIP---VTMANA-----769
Db 1311 VTMDNILSG--FENEYDVYIK-PL-----AGVYSLKQIEKNIITFNJLNDILNSRLK 1363
QY 770 ----YLD-----NOSTYIVVEPILEKENQNDKPSILPOFRNK-----A 804
Db 1364 KRKYFLDVLSDLMQFKHISSNEVIE--DSFKLLNSEQNTLLKSKYKIKESVENDIKFA 1422
QY 805 QENSKLDEKV-----EETSEKVEKEKLSLSTGNSSTLEEYPTVDPQVEKVAKF 856
Db 1423 QEGISYKVKVLAUKDDLESIKKVIKEKEKPSPTTPPS-----PAKTDEQRKESKF 1477
QY 857 A-----ESYGMKLENVLFNM-----DGTIELYL 879
Db 1478 LPFLNIETLNNLVNKIDDDVILNKLAKINDCNVEKDEAHVKITKLSLKAIDOKIDLF- 1536
QY 880 PSGEVYKKNMADFTGEAPQNGENKPSENGK-VSTGIVENQPT-----ENKPADSLPE 931
Db 1537 -----KNTNDFRAIKKLINDDTKMDLGLLSTGLVQNFPTIISKLIEGKFDMLNI 1589
QY 932 APNE--KPVKPPENS 943
Db 1590 SOHCQVKKQCPENS 1603

RESULT 3
CNA.STAAU
ID CNA.STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoesek M.;
RT "Molecular characterization and expression of a gene encoding a
Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoesek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoesek M.;
RT "Identification and biochemical characterization of the ligand
binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RP MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
Moore D., Jin L., Schneider A., DeLucas L.J., Hoesek M.,
Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
aureus adhesin.";

```

```

RL  Nat. Struct. Biol. 4:833-838(1997).
CC  -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC  COLLAGEN-CONTAINING SUBSTRATA.
CC  -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC  an amide bond (Potential).
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
DR  EMBL; M81736; AAA20874.1; .
DR  PDB; 1AMX; 24-JUN-98.
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  TIGRFAWS; TIGR01167; LPXG_anchor; 1.
DR  PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
KW  Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT  SIGNAL 1 29
FT  CHAIN 30 1154
FT  PROPEP 1155 1183
FT  DOMAIN 151 318
FT  DOMAIN 533 1093
FT  DOMAIN 1093 1157
FT  REPEAT 533 719
FT  REPEAT 720 906
FT  REPEAT 907 1093
FT  SITE 1151 1155
FT  MOD_RES 1154 1154
FT  SEQUENCE 1183 AA; 133066 MW; 6A1CC072E575D76 CRC64;
SQ
Query Match 3.9%; Score 205.5; DB 1; Length 1183;
Best Local Similarity 19.7%; Pred. No. 0.027;
Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;
QY 16 NNRVSVVDGSSQSKSENLT-PDOVSQKEIQAEQIVKIKITDQGYVTSHGDHYHYNGK- 73
DB 155 NNEKSV-----SKDITIKDQGGQOLDLSTLINWT-----GTHSNVYSGS 238
QY 74 --VPYDALFSEELMKDPNYQLKADI-----VNEVKGGYIYIKVDG---KYVYLKDA 122
DB 239 AITDFEAPGSKITVDNTKNTIDVTIPOQYSGNSFSINYKTKITNEOQKEFVNNSQAW 298
QY 123 HADNRTKDEINQKQEHVKDNKENVN-----AVARSQGRYTIN 163
DB 299 YQEH--GKEEVNGKSFNHTVHNINAGIEGTGVKELKVKDKDTKAPIANVKFKLSKK 356
QY 164 DGXYF----NPADIIEDTGNAYIVPHGGHYHYIPKPSDLSASELAAKAHLAGKNMPSQL 219
DB 357 DGSVVRDQKEIITDANGI-----ANIKALPSGDYILKEAPRPYTFDKDKE---- 406
QY 220 SYSTASDNTQSVAGSKSPANKSENQSLKELYDPSAQRYSESGLYFDPKAIIS 279
DB 407 -YPFTKRTDQGYF--TTTENAKAEKTKDV-----SACK-----VWEGTQKVK 448
QY 280 RTPNGVAIPGHDHYHFIYPSK--LSALEEKIARMVPISGTGSTVSTNAKPNVSVLSGL 337
DB 449 PTYFKLYKQDDNQNTTPVDKAIKKLED-----GTTKVTWNLSPND----- 491
QY 338 SSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYTPKS-NQIQDPTLP- 395
DB 492 -----KNGKAIKYLKVEVNAQGEDTTP 514
QY 396 -----NNSLATPSPLPI-NPCTSHKHEHDGFGFDANRIIADESGFVMSHGDHNYHF 449
DB 515 GYTKKGLVVTNTEKPIETTSISGEKVDKDNQDKR--PKVSVNLLANG----- 565
QY 450 KKDLEBQIKAAQKHBEVKTSHNGDLSHSEQDYPGNKEMKDKDKKIE-EKIAGIKM 508
DB 566 -----EKVKT-----LDVTSETNWKY-----EFKDLPKYDEGKKIEYTVT 600

```

```

QY 509 QYGVK-----RESIVVNK-----EKNAILYPHGDHHHADDIDEHKPGVIGSHSNVELEK 558
DB 601 EDHVKYDVTDTINGTTITNKYTPGTSAITVKNWDDNNQ--DGKRPTET-----KYVELY- 652
QY 559 PEEGVA-----KKEGNKV-YTGEELTNVYNLLKNSTFNNOFET 595
DB 653 -QDGKATGKTALINESNNWTHWTGIDEKAKGOQVXYVEELTKV---KGYTHVDND 707
QY 596 LANGQRKVSFPPELEKKGILNKLITPDGKVLKESGKVGEG-----VGNIAN 648
DB 708 M--GNLIVTNKYTPETTSISGEKVDKDNQDKRPEKSVNLLADGKVKTLDTSETN 765
QY 649 FELDQPLP---QGTFRKYTIAS---KDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAG 701
DB 766 WKYEFKDLPKYDEGKKIEYTVTHDVHVDY--TTDINGT-----TITNKYTPG 810
QY 702 DTYLRV-----NPQFAVPKGTDALRVFDE-----FHGNAYLENNKYV 739
DB 811 ETSATVTKNWDNNQDKRPTETKVELYQDGKATGKTALINESNNWTHWTGLDEKAG 870
QY 740 GEIKLPIPKLNQ---GTRTAGNKIPVTFMANAYLDNQSYIVVEPILKEKNOTDK--- 792
DB 871 QQVAYTVEELTKVGYTHVDNDMGNLIVTNKYTPETTSISGEKVDKDNQDKRPEK 930
QY 793 -PSILPQFKRKA-----QENSKLDE--KVEEPTKSEKVEKLESETGNST 835
DB 931 VSVNLLANGEKVKTLDTSETNWKYEFKDLPKYDEGKKIEYTVTHDVHVDYDITDINGTTI 990
QY 836 SNSTL--EEVPTVD-----PVEKVAKFAESYGMKLENVLFNMDGTIELYLP 880
DB 991 TNKYTPGETSATVTKNWDNNQDKRPTETKVELYQDGKATGKTALINESNNWTHWTG 1050
QY 881 SGEVTKKNMADFT-GEAPOGNENKPSKSVGTGTVENQPTENKPADSL-PEAPNEK-- 936
DB 1051 LDEKAKGOQVXYTVDELTKVNGYTHVDNDMGNLIVTNKYTPKPKPKPIYEPKPKDTP 1110
QY 937 PVKPENS 943
DB 1111 PTKPDHS 1117
RESULT 4
IGA4_HAEIN
ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH71 HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulson K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE.
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).

```

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M87491; AAA24968.1; .
 DR MEROPS: S06.001; .
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertactin_sup.
 DR Pfam: PF02395; IGA1; 1
 DR Pfam: PF03212; Pertactin; 2.
 DR PRINTS: PRO0921; IGASERPTASE.
 KW Hydroxylase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1021
 FT PROPEP 1022 1849
 FT ACT_SITE 299
 FT SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;
 SQ
 Query Match 3.8%; Score 203.5; DB 1; Length 1849;
 Best Local Similarity 19.3%; Pred. No. 0.062;
 Matches 239; Conservative 153; Mismatches 449; Indels 397; Gaps 59;
 QY 8 HRSQENKNNRYSYVDGSSQSKSENLTLDQVSKQEGIQ-----AEQIVIKITDQ 57
 DB 125 HRDVSSENR---YTVKNNPFTEVNTSFTTKBQDQARRDYMPPLDFVTEVAPI 181
 QY 58 GYVTSIGDHYHYNGKVPYDALF-----SEELMKDPNQOL-----KDAD-----IVNEVAGG 105
 DB 182 EASTANNKGE-YNNSDKYPAFVRLGSGSQFIYKKGSRVQLILTEKDKGNLLRNWDVGG 240
 QY 106 YIKVDGKYVYVLKAAHADNVRTDEINRQKHVKNEKNSVNAVARSQRTTNDG 165
 DB 241 DNLELVGNAYTY-----GLAGTPYKVNHEN-----NGLIGFNSKEEHSDFKG 283
 QY 166 YV-----FNPADIETDGNAYIV--PHGGHYHYIPKSDL-----SASELAAKAAHLAKG 212
 DB 284 ILSQDPLTNVAVLGDGSGFLFVVDREKGNWFLGSDFWAGYKKSQWENNYYKHEPAEK 343
 QY 213 NNQPSOLSYSS-TASDNTQSV--AKGSTKSPANKSENLSQSLKELYSPSAQRY-----264
 DB 344 IYQ-----QYSAGSLTGSNFTQYTWQATGSTTITGGCEPLSVDLTDGKDPKPNHKGKSIILKG 399
 QY 265 -----SESQGLVFDPAKIIISRPN-----GVAIPHGD-----HYHFIPYSKLS 302
 DB 400 SGTLLNHNIDQAGGLFPEGDYEVKGTSDSTTWKAGVSVADGKTVTKVHNPKYDRLA 459
 QY 303 -----ALEEK-----IARMVPISGTGSTVSTNAK- 326
 DB 460 KICKGTLVVEGKGNELGLKVGDTGTVLKQKADANKVQAFSQGVISGRSLVLNDDQK 519
 QY 327 --PNEVVSLL--GSLSSNPSSL-----TTSKELSSADGYIFNPKD 363
 DB 520 VDPNSYIFGFRGRDLNGLNSLTFDHIHNDGCAVNVHNMNTNIIITGSLTNP-- 577
 QY 364 IVEETATAVIVRHGDHFH-----YIPKSNQIQGQTLPLNNSLATPSPSLINPG 411
 DB 578 ---NTITSYNIEAQDDHPLRIRSPYQLYFNQDN--SYVTLKGG--ASTSESLPQNSG 631
 QY 412 TSHKHEEDGYGFD-----ANRIIADESQGVMSGHGDNHYFFKDKDTEBOIKAAQKHL 465
 DB 632 ESENWNLVNGRTSDPAKRVNMHINERNNGF-----NGYF-----GEETKATQN-- 677
 QY 466 ESKVTSNGLDLSLSHEQDY--PGNAKEMKDLDDKIEEKIAGIMQYGVKRESIVVNKEK 523
 DB 678 GKLNVTENG-----KSDQNFLLTGTNLNGDL-----NVKEK 709
 QY 524 NAIITPHGDHHDHADIIDEKHPVGIG-----HSHSNYELFKPEEGVAKK-----EGN 569

DB 710 GTLFLSGRPTPHARDI-----AGISSTKKDHPHTENNEVVVEDDWINRNFKATTMVGTN 764
 QY 570 -KVYTGELTNVNVLLKNSTNNQNFILANGQKRVSFSPPELKKIGLINMLVK-----622
 DB 765 ASLYSGR--NVANITSNITASN-----NAQVILGY-----KUGDTVCVRSDDITGY 807
 QY 623 LITPDCKVLEKSVKVGEGV-GNI-----ANFELDQYLPQGFYKTYTIASKDYPEVSYD 676
 DB 808 VTCNSNLSKALSNFNTNLRGNVNLTEASFTLGKANLFG-----TIQSIGTSQVN-- 860
 QY 677 GTFTVPTSLAYKMASQTIFFPHAGDTVLRVNPQFAVPKGTDALVRVDFEHGNA-----731
 DB 861 ---LKENSHWHLTGNVNVNQLNLITNGHILHNAQDANKVTTYNTLVNSLSGNGSFFYW 916
 QY 732 -----YLENNYKVGEEKLPI-----PKLNQGT-----TRTAGNKIPVTMANAYL 771
 DB 917 VDFTNKNSKVVVNAKSGATGNFTLOVADKGTGEPNHNELTLDASNATRNLEVT-LANGSV 975
 QY 772 DN-----QSTYIIVEVILEKENOT-----DKPSILPQFKNKAQEN 807
 DB 976 DRGAWKYKLRNVNGRYDLYNPEVEKRNQTVDTTNTITPNDIQADAPS-----RQSN 1026
 QY 808 SKLDEKVEEP-----KTSEKVEKEKLSETGNST 835
 DB 1027 NEETARVETPVPPAPAPATESAIASEQPETRPAETAQAQPAEMEETNANSTETAPKSDTATQT 1086
 QY 836 SNSITLEEVPTVDYQKQVAKFAESYGMKLENVFNMDGTIELYLPSEVYKKNMADF---892
 DB 1087 ENPNSSEVPS-----ETTEKVAENPPQENETVAKNEOEATEPTQNGEVAKEADQPTVEAN 1141
 QY 893 --TGAPGCGENKPSKNGKVTGTGVNQPT-----ENKPADSLPPAPNPKPV-----K 939
 DB 1142 TOTNEATQSECKTEETQ-----TAETKSEPTSVTSVENQPEKTVSQSTEDKVVVKEEK 1196
 QY 940 PENSTDNGLMPLNEGVNGSDP-MLDPALPEEAPVDPVQE 976
 DB 1197 AKVETEETQKAPQVTSKPEPPKQAPAEPAPEEVTDTNAEE 1234
 RESULT 5
 MSP1_PLAFC
 ID MSP1_PLAFC STANDARD; PRT: 1726 AA.
 AC P04934;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAR-1989 (Rel. 10, last sequence update)
 DT 01-OCT-1996 (Rel. 34, last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5835;
 RN (1)
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 RL the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 CC (2)
 CC SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=98143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolif R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 RL malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/send an email to license@sb-sib.ch>).

EMBL: X03831; CAA27446.1; -	P.T.R: A23386; SRZOGM.	InterPro: IPR000561; EGF-like.	Pfam: PF00008; EGF; 1.	Malaria; Merozoite; Polyprotein; Repeat; signal; Glycoprotein; Transmembrane; GPI-anchor.	POTENTIAL.
1	19	SIGNAL			
2	1726	CHAIN			MEROZOITE SURFACE PROTEIN 1.
3	133	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
4	272	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
5	501	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
6	567	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
7	638	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
8	827	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
9	924	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
10	944	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
11	990	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
12	1016	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
13	1114	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
14	1221	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
15	1613	CARBOHYD			N-LINKED (GLCNAC. . .) (POTENTIAL).
16	1736	SEQUENCE			D80AD45FA352BCF3 CR664.

1021	--NDDIKQFVKSNSKVITGLTETOKNALNDEIKKLKDTQLQSLFYKNRYKLLKDLRFNKK	1070
483	QDYPGNAREMKD---LDKKIEKIAGIMQKGVKRE--SIVNKEKNAIL-----	527
1079	KEIGQDKWQIKKLTLLKEQLSKLSLNPHNVLQNFVSFFNKKEAEIAETENTLENTK	1138
528	-----YPHGDHHADPIDEHKPVGI--CHSHSNVELFKPEGVAKKGNKVYTG	574
1139	ILKHVGLVXYNGE---SSPLKTLSEVSIQTEDNVANLEKFEVLSDIKDGLNHLHG	1195
575	EE-----LTNVYLLNKKSTFNQNTFLANGOKRVSFSPFPPELEKGLINMLVKL---	623
1196	KKKLSFLSSGLHHLITELK-EVIKNKNT-----GNSPSENKK--VNEALASKYEN	1243
624	IFPDGKYLEVSKGVFGEVGNIANFELDQYLPQGTFKTYI-----ASKDYPEVSDGT	678
1244	FLPEAKVTVTVIP-----PQDVPSPLSVRVSGSGSTKEETQITPSG-	1287
679	FTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVRPGTD-----ALYRVFDEEH	728
1288	-----SLTLEQQVVLQNYDEEDDSLVLQIFGESEDNDEYLDQVVTGEAISVTWGNIL	1342
729	GNAYLENKYKGEIKLPIKLNQGTTFTAGNKIP--VTFMANA-----YLD	772
1343	SG--FENEYDVYTLK-PL----AGVRSKKQIKENITFTNLNLDLSRLKRYFLD-	1395
773	-----NOSTYIVEPILKEKNQTDKPSILPOFKRNK-----AQENSKLD	811
1396	VLESOLMOPKHISSENYTIE-DSEKLLNSEQKNTLLASYKYIKESVENDIKFAQEGISYV	1454
812	EKV-----EETPKTSKVEKEKLSGNTSNTLEEYPTVPQEVAKFA-----	857
1455	EKVLAKYKDDLESIKVKIKEKEKFPSPPTTPS-----PAKTDEKKESKFLPLTNI	1509
858	-----ESYGMKLENLFNM-----DGTIELYLPSEGVK	886
1510	ETLYNLYNKIIDYLYNLKAKINDCNVEKDEAHVKITKLSLKAIDKIDLF-----	1561
887	KNMADFTGEAPGNGENKPSNGK-VSTGTVENOPT-----ENKPADSLPEAPNE--K	936
1562	KHNDEFAIKKLINDTKDMLGKLLSTGLVQNFPTNIIKLTGKFDQMLNLSHQCVK	1621
937	PVKPENS 943	
1622	KCPENS 1628	
RESULT 6		
MSPL_PIAPP	MSPL_PIAPP	
ID	MSPI_PIAPP	STANDARD: PRT: 1726 AA.
AC	P50495;	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DE	01-OCT-1996 (Rel. 34, Last annotation update)	
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)	
DE	(PMSA) (GP195).	
GN	MSP-1.	
OS	Plasmodium falciparum (isolate palo Alto / Uganda).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OC	NCBI_TaxID=57270;	
RN	[1]	
RR	SEQUENCE FROM N.A.	
RR	MEDLINE=89005525; PubMed=3049134;	
RR	Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,	
RR	Siddiqui W.A.;	
RA	"Plasmodium falciparum: gene structure and hydropathy profile of the	
RT	major merozoite surface antigen (gp195) of the Uganda-Palo Alto	
RT	isolate." RT	
RT	Exp. Parasitol. 67:1-11(1988).	
CC	!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor	
CC	(Potential).	
CC	!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42	
CC	kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF	

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M37213; AA29611.1; --
 CC InterPro: IPR000361; EGF-like.
 CC Pfam: PF00008; EGF; 1.
 CC Malaria: Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 KW SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 1726 AA; 196174 MW; 5B59CEBFA2F9A026 CRC64;
 CC
 CC Query Match 3.7%; Score 199; DB 1; Length 1726;
 CC Best Local Similarity 19.7%; Pred. No. 0.091;
 CC Matches 239; Conservative 168; Mismatches 402; Indels 402; Gaps 63;
 CC
 CC 3 YALNHRSEN---KDNRFVYVDGSSQSKSNTLPDQVSKQEGIQAEQIVIKITDQG 58
 CC 550 YYYNEKQYNNKFFSSNNVNV---OKLKAALSYLED-YSLRGI----- 592
 CC 59 YVTHGCHVHYNGKVPYDA---LSEEL-----LMKDPNYQLKADIVN- 100
 CC 593 ---SEKDFNHYTLTGLEADIKKLTETEEKSSKENKLEKNFKGLTHSANASLEVYDVKL 649
 CC 101 EVKGGVII-KVDG--KYVYVLKADAAHADNVRT-----XDEINRQKQ- 138
 CC 650 QVQKVLIIKKIEDLKIELFLANQKLSIHVPNIYKQNKPEPYIYLVLRKEVDKLEF 709
 CC 139 ----EHVKDNEKVNVA---VARSQRYTTNDGIYVFNPAIIEDTGNAYIVPRGGHYH 190
 CC 710 IPKVDMKKKEQAVLSITQPLVAASE---TTEDG-----GHSTH 746
 CC 191 YIPKS-----DLASELAAKAHLAKGNMOPQSLSYSTA---SDNNTQSVAKGSTS 239
 CC 747 TUSQGETEVTETETETVGHHTTITLPPKVKVYVENSIEHKSNDNSQALKTIVYL 806
 CC 240 KPANK-----SENLOSLLKLYD-SPSAQRYSESDGLVDFDPAKIISRTPN 283
 CC 807 KKLDELFTKSYCHKYILVSNSSMPQKLLVYNLTPPEENELKS---CDPLDLILFNQN 862
 CC 284 GVAIPHG-----DHYHFTPYSK-----LSALEEK-----TARMVPTSGTGSTVST 323
 CC 863 NTPAMYSLDSYMNNDLQHLFFELYQKEMIIYLLHKLKEENHKKLLEEQKQITGTSTSSP 922
 CC 324 -NAKPNVYSSLSGSSNPSSLTTSKELSSASDGYIFNPKDIVEATATAYIVRHGDHRYH 382
 CC 923 GNTVNTAQSAHNSNQNSQNSASS---TNTONG-----VAVSGGAVVEESHDPLTV 972
 CC 383 IPKSNQI-QGPTLPN--NSLATPSPSLPINTPGTSHKHEEDQYGFDAIRIAEDESFGVM 439
 CC 973 LSTNSDLKGIIVSLLNGKTKVNP-LTIST-TEMEKFFEN-----ILKNNDTYF-- 1020
 CC 440 SHGDHNIHFFPKD-----LREQIKAAQKHLEEVKTS-----HN-----GLDSLSSHE 482

Db 1021 --NDIKQFVSNKSVITGLTETQKNALNDIEKKLKDQLSFLINKYKLDRLFNKK 1078
 Qy 483 QDYPGNAXEMKD---LDKKIEEKIAGIMKQYGVKRE-SIVVNKEKNAIL----- 527
 Db 1079 KELGQDKMQIKLTKLLKQLESKLNSLNPHNVQLNESVFFNKKKKEAIEAETENTLTK 1138
 Qy 528 -----YPGCHDHHADPIDEHKPVGI--GHSHSNYELFKPEPGVAKKEGNKYVTG 574
 Db 1139 ILLKHYKGLVYYNGE---SSPLTKLSEVSIQTEDNANLEKFRVLKIDGKLNDNLHLG 1195
 Qy 575 BE-----LTNVNLLKNSTFNNQNTFLANGQKRVSFSPFPELEKKLGINMLVK 622
 Db 1196 KKKLSFLSSGLHQLITELKEVINKNKNYTGNS-----PSENKKK--VNEALK 1239
 Qy 623 L---ITPGCKVLEKVGKVGEGVGNANTANFELDQPLPGQTFKYTI-----ASKDYPEVS 674
 Db 1240 SYENFLPEAKVTTVVTP-----POPDVTIPSLSVRSVSGSGSGSTKEETQIP 1284
 Qy 675 YDGTFTVPTSLAYMASQTIIFYPPHAGDTYLRVNPQFAVPKGTD-----ALVRVF 724
 Db 1285 TSG-----SLTLELQVQVQLQNYDEEDDSLVVLPIFGESEDNDEYLDQVVTGEAIVTM 1338
 Qy 725 DEFHGNAYLENNYKVEIKLIPKLNQGTTRTAGNKIP---VTMANA----- 769
 Db 1339 DNILSG--FENEYDVIYK-PL---AGVYRSLKKQIEKNIFTENLNLINDILNSRLKRR 1391
 Qy 770 YLD-----NOSTYIVVPILEKQNDKPSILPQFKRNK-----AQEN 807
 Db 1392 YFLDLESDLQKQFHISSNEVILIE-DSFKLLNSQKNTLLKSYKIKESVENDIKFAQEG 1450
 Qy 808 SKLDEKV-----EPEKTSKVEKEKLSGTSTNSTLEEVYTPDVPQEKVAKFA-- 857
 Db 1451 ISYVEKVLAKYKDDLESTKKVKEKEKFPSPPTTPS-----PAKTDEOKKESKELPF 1505
 Qy 858 -----ESYGMKLENVLFNM-----DGTIELYLPSPG 882
 Db 1506 LTNITLYNNLVNKDILYLLNLKAKINDCNVKEDEAHVKITPKLSLDLRAIDDKIDLF 1561
 Qy 883 EVIKKNMADFTGEAPQNGENKPSNGK-VSTGTVENQPT-----ENKPADSLPEAPN 934
 Db 1562 ---KNHNDFAIRKLINDTFKDMLGKLLSTGLVQNFNPNTIISKLIEGRFQDMNLISQH 1617
 Qy 935 E--KPVKPENS 943
 Db 1618 QCVKRCQCPENS 1628
 RESULT 7
 MRSP_STAAU STANDARD; PRT; 1637 AA.
 ID MRSP_STAAU STANDARD; PRT; 1637 AA.
 AC P80544; Q92F62;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methicillin-resistant surface protein precursor.
 GN PLS.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate 1061;
 RA Savolainen K., Kuusela P., Paullin L., Korhonen T.K.,
 RT "Pls, a large repeat-rich surface protein of methicillin resistant
 Staphylococcus aureus";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
 RP 1199-1205 AND 1217-1224.
 RC STRAIN=Isolate 1061;
 RA MEDLINE=96270743; PubMed=8665912;
 RX Hilden P., Savolainen K., Tyyneä J., Vuontola M., Kuusela P.;


```
CC -----
DR EMBL; 246791; CA86762.2;
DR EMBL; 246792; CA86762.2; JOINED.
DR EMBL; 246792; CA86767.2;
DR EMBL; 246791; CA86767.2; JOINED.
DR WormPep; C09G5.8; C025774.
KW Hypothetical protein.
FT DOMAIN 384 394 POLY-SER.
FT DOMAIN 395 398 POLY-ASN.
FT DOMAIN 1150 1154 POLY-ASP.
SQ SEQUENCE 1531 AA; 174862 MW; A453990AF0DEBC2F CRC64;

Query Match
Best Local Similarity 3.6%; Score 189.5; DB 1; Length 1531;
Matches 197; Conservative 160; Mismatches 441; Indels 229; Gaps 39;

QY 113 KYYVYLKRAHADNVTDEINRQKHVKON-----EKVNSNVAVARSQGYTTNDGYV 167
DB 170 KYNLRDRTSSSDRYARSARYPIEKWRSPOLDEHFHNVVELNAKQAKKVEQEKQITFNRSR 229
QY 168 FNPADIEDTGNAVIVPHGGHYHYPKSDLSASALAAGAKAHLAGNMOPSOVSYSTASD 227
DB 230 FRSMLEKRSQNEKVVERSKYDDVKEQNLDMKLAQAQQL-----LIYTPASAR 280
QY 228 NNTQSVAKGSTS-----KPAKSENLOSLLKELYDSPAQYS--ESDGLVFDPA 275
DB 281 ATTASMTGRSTFROPPSTFRQRPPLTAGTGS1-----DRPGSAPVARKKSDG---GE 331
QY 276 KIISTPNCVAIPHGDHYHIFYSKLSALEAKIARWPISTGTSVSTNAKPNVWVSLG 335
DB 332 KLOATDEKLAIVRLNRLKNDIEITELKYIEKLRLK-----SSVNOSSPTRLSTSS 388
QY 336 SLSSNPSSLTTS-----KELSSADSGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQ 388
DB 389 SKSSSSNNNDGEGKDSLEENSESDSGRSTPVIEKKKPR-RKSRKSSHQPSKNP 447
QY 389 IQOPLPNNSLATPSPPIPGTSEKHEEDCYCFDANRIIAEDSEGVSHGDHNYIF 448
DB 448 IPPPRIPDQTEKVLKDLKVAENDLAMIQECDLYKANERLVHOSLSKSTEYGAESIE 507
QY 449 FRKDLT--EEQKAAQKHLVEE---KTSNGLDLSLSEHQDYPGNAKEMKDLKKIEEK 502
DB 508 EKKKIVLEELKETEKRKESEHRRERDQKKEAMRLH-----YKNKYDAKKTEKK 560
QY 503 IAGIMQYGVKRESIVNKEKNAIYPHGDHHDPIDEHKPVGIGHSHNVLEPKPEG 562
DB 561 LSVAKNSKVEEERT--EEKIS-----HSPPTMTFPIRRHQSQSEISRMRADD 609
QY 563 VAKKECNKYVTGEELTNV-----VNLLKNSTFNQNFILANGQKRVSPFPELEKKLGI 617
DB 610 DILLQ---KLY--KEVADILHSDVGIAINTLGASENSLAKWQKLYSELY-EELEKVR-- 661
QY 618 NML-----VKLITPDGKVLKVGKGVGEGVGNIANFELDPQYLPQGTFKYT 664
DB 662 NMLLIQYDINQOMKEIKLLKDELRLKTVSAEILSK-----SREVEE-----PQKKIF 711
QY 665 IASKDYPEVSDG-----IFTVPTSLAYKM-----ASQIFYPFHAGD 702
DB 712 MLEEQIRTIAYSGQPKVLLANQIMPIPRVNTDLSVLKINLVKPSPTSSTFFFSLEFFD 771
QY 703 TYLRVNPQFAVFKGTDALVRVDEPHGNA---YLENNYKVGVEIKLPIPKLNGQTRTAGN 759
DB 772 FQLEITTPMDAKQHNDFTTVYDVLVSNLLIHYLQTNQIVIEVMPASDCYKLLAAATIS 831
QY 760 KIPV-----TFMANAYLDNQSTVIVEVPILEKENQTKDPSILDPQFRKNRAQENSK--- 809
DB 832 LIPLFEDSVLRFCSIEMLKSDYTG-VMCTLRVIEVSQP-ISDSFKFKKSEMARNNL 889
QY 810-----LLEKVEPEKTSSEKVEKEKLSSENGSTNSTLFE- 842
DB 890 PLOLENEDETDNFDPLIMVNRVVGLTDFGKDPSTFCIVDEFISFPYPTDFTSISSEI 949
QY 843-----VPTDPVQEKVAKFAESYGMKLENVLFNMDGTI-ELYLPSGEVIKKN----- 888
```

```
DB 950 RSKRDCYIPKIDIIARNLFATSSISFFEL-IENIP-RQGVIAITLHLPLHPLCKLGSGIKGT 1007
QY 889 --MADFTG-----EAPQNGENKPSSENGK-----VS 912
DB 1008 FPLMDTGRSPSSVSLDLCLINKHEIPSPFFLKHEPKELKEVKVDTPILPQPVRRTSKEFVV 1067
QY 913 TGTVENQPTENKPADSLPEAPNEKVPK-----ENSTDNGMLNPEGVNGSDPMLDPALE- 966
DB 1068 TPVKEAELHDAEPTSMPPKAP-EPTTAPLRRLSTDSSTSFSSKSDLFSPTNPQTYDY 1126
QY 967 EAPAVDP 973
DB 1127 EIPAVTP 1133

RESULT 9
P3K1_DICDI STANDARD; PRT; 1570 AA.
AC P54673;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Phosphatidylinositol 3-kinase 1 (EC 2.7.1.137) (PI3-kinase)
DE (PtdIns-3-kinase) (PI3K).
CN PI3A OR PIK1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBL_taxid:44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT discoidium: biological roles of putative mammalian p110 and yeast
RT Vps34p PI 3-kinase homologs during growth and development.;
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -! CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-LD-myo-inositol = ADP +
CC 1-phosphatidyl-LD-myo-inositol 3-phosphate.
CC -! SIMILARITY: BELONGS TO THE PI3/PI4-KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U23476; AAA85721.1;
CC DictyDb; DD01099; p1Ka.
CC InterPro; IPR002420; PI3K_C2.
CC InterPro; IPR000341; PI3K_ras_bind.
CC InterPro; IPR001263; PI3Ka.
CC InterPro; IPR000403; PI3_P14_kinase.
CC InterPro; IPR000626; Ubiquitin.
CC Pfam; PF00240; ubiquitin; 1.
CC Pfam; PF00454; PI3_P14_kinase; 1.
CC Pfam; PF00613; PI3Ka; 1.
CC Pfam; PF00792; PI3K_C2; 1.
CC Pfam; PF00794; PI3K_rbd; 1.
CC SMART; SM00142; PI3K_C2; 1.
CC SMART; SM00144; PI3K_rbd; 1.
CC SMART; SM00145; PI3Ka; 1.
CC SMART; SM00146; PI3K; 1.
CC PROSITE; PS00915; PI3_4_KINASE_1; 1.
CC PROSITE; PS00916; PI3_4_KINASE_2; 1.
CC PROSITE; PS0290; PI3_4_KINASE_3; 1.
CC Transferrase; Kinase; Multigene family.
KW DOMAIN 41 48 POLY-ASN.
FT DOMAIN 59 73 POLY-ASN.
FT DOMAIN 161 168 POLY-ASN.
```

FT DOMAIN 294 304 POLY-THR.
 FT DOMAIN 308 324 POLY-THR.
 FT DOMAIN 413 424 POLY-ASN.
 FT DOMAIN 503 510 POLY-SER.
 FT DOMAIN 570 579 POLY-ASN.
 FT DOMAIN 821 828 POLY-ASN.
 FT DOMAIN 831 836 POLY-ASN.
 FT DOMAIN 1309 1570 P13K/P14K.
 SQ SEQUENCE 1570 AA; 178374 MW; 558678B72B34D783 CRC64;

Query Match 3.58; Score 186.5; DB 1; Length 1570;
 Best Local Similarity 17.68; Pred. No. 0.3; Indels 461; Gaps 48;
 Matches 205; Conservative 141; Mismatches 361;

QY 128 RTKDEINQKOE--HVKDEKV-----NSNVAVARSOGRYTTNDGYVFNPAADIIEDTIGNA 180
 DB 77 KRKDSKQNGEINQEMSENKKIYNDSNCSGSSGGHVNNGHHI-----LIEEN--- 128
 QY 181 YIVPHGGHYHIPKSDLSASELAAKAHLAGKNQPSOLSSTASDNNTOSVAKGSTS- 239
 DB 129 -----ERLEHENQEIQIYKOK---GMEFQKDLREGYDVNSNNNNNGGSSG 175
 QY 240 -----KPAKSEN-----LQSLKEL-----YDSPSAQRYSES 267
 DB 176 SSSGSDSASNOPIRTNREGSILNKKOGLYKEISQRFQTPDTASYTRPNANNISIK 235
 QY 268 DGLVF-----DPAKIISRTPNGVAIPHGDHYHIFPYKSLSALEEKIAR 310
 DB 236 DKISILKREQERRKQDSEVOQREKIVIVLSADSNIIQIYHPVSLIEKMKSKLDTTEKPATT 295
 QY 311 MYPISGTSTVSTNAKPNVYSSIGLSNPSL-----TTSKELSSADGVIFNPKDIVE 366
 DB 296 TTTTITSTISSTPTTTTITTTTSTNDITIKPKTSPTKNNEERSQSPLITPKQPV 355
 QY 367 ETATAYIVRHGDHFHYIPKSNQIOGPTLPNLSLATPSPSLINPQTSHEKHEEDCYGDA 426
 DB 356 E-----IVKKVS-----TPKSN-----NTSKTSSDTPTGKTKTKDKDKDKSRDS 398
 QY 427 NRIIADESGFVMSGDHNN-----VF----- 448
 DB 399 GNLVIYVNTNTSSNNNNNNNNNNETIKRRGRVLVTPSPDLKKNQIYFTIPINPV 458
 QY 449 -----FKKDLTEEQI-----KAAQK 463
 DB 459 NKTNKNQLLNNTSQOFLKTLNISNIPDCKINDINDTDAFSDLSASASSSFIKSSQS 518
 QY 464 -----HLEVKTSNGLDLSLSSHEQDYPGNAKEM-----KDLKKIEEKIAGIMK--- 508
 DB 519 LLNVQSLRVKAKTSFNILFLM-----PNOSKILQVKGSDTIENLKERIISDYLFNN 571
 QY 509 -----QYGV-----KRESIVYVK-----EKNAIY 528
 DB 572 NSNNNNNCKYGADSYLLIFDNDNPMERSVLNKSQYILDKRAQGLPKLKVIERKSTIL- 630
 QY 529 PHGDHHAIDIDHKVPGVGHSHSNVELEK-----PREGVAKKE----- 567
 DB 631 -----DSDPDSLEP-----SEYIIRKLIPGDTWRGEEVEYFRVTRRLRYEALP 677
 QY 568 -----GNKYVTG-----BELTNVWLLKNST-----FNNQNFITIA 597
 DB 678 LKGSQSTILLVPLSPPIPVGNKLLISLIFLPIQVTKTLDLEUNETADQFTNRLFT-- 735
 QY 598 NGQKRVSFPPPELEKLGINMLVLIITPDGKVLKSVKGVGEGVGNANTANE-LDQVYL 656
 DB 736 ---KNYSKHLN-----VSNHFTLVKVGSSDFIHGPHDTRTFESIRNHII 778
 QY 657 PGQTFXYTISKDYPVSVYDGTFTVTSLAYKMASQTIYPFHACGDTYLRVNPQFAVPGK 716
 DB 779 QGKTPQTLTQRKPELDPO-----PP-----KPRDYPPE 809
 QY 717 TDALVRVFDEFGHNAYLENN-----YKVGKILPIPKLNOGTR 755
 DB 810 L-----IIDHSCSNAINCNCNNNTSTNNNNINFDNDWDQITHISIREIKKPFVKVMGSTR 864

QY 756 TAGNKTPVTFMANAYLDNQSTVIVEVPILK-----ENQTDKPSILPQPK----- 800
 DB 865 -----IPLSCIKDI-----DSSSVIVSISLYHGIECFSKAFTQIIPPPFAFLAETLSVDMC 916
 QY 801 -----INKAQENSKLDEKVEEPTSEKVEKLSERGTNSTSTLEEVPTVDPVQEKVAK 855
 DB 917 EWLVFTNIDYSNLPVDARL-----SISVYSANETVDDVE----- 951
 QY 856 FABSYGKMLNVLNFMGDTIELYLPSPGEVIKKNMADFTGEAPQNGENK--PSENGKYST 913
 DB 952 -----IKNLDEATKKLTPIG-WINVMITDFKYQLRQGMVLSLWPSDFS-N-PL 997
 QY 914 GTVENOPTENKPADSLPE-----APNEKPKVPKPSNSTDNGML-NPEGNVGSDFMLPALLEE 967
 DB 998 GTCNNPSSSQSVGLTLEFEFNLPVLFRKTKFSTSVSIEQPTNINSNEMRE-FTFQ 1056
 QY 968 APAVDVPQOE-KLEKFTASYGLGLDSVIF 994
 DB 1057 ITALDPLSLDLKQEKYNQLWTLRHSYILF 1084

RESULT 10
 SC16_YEAST STANDARD; PRT: 2195 AA.
 ID SC16_YEAST Q02822;
 AC P48415; Q02822;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 13-JUN-2002 (Rel. 41, Last annotation update)
 DE Multidomain vesicle coat protein.
 GN SEC16 OR YPL085W OR YPL1W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96017704; PubMed=7593161;
 RA Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
 RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that
 RL interacts with Sec23p.";
 RL J. Cell Biol. 131:311-324(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,
 RA Araulo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Drells H., DiPaolo T., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
 RL Nature 387:103-105(1997).
 CC -!- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE
 CC ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
 CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
 CC CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION
 CC CAUSES A LETHAL SECRETION DEFECT.
 CC -!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
 CC WHICH BUD FROM IT.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; U23819; AAC49088.1; -
CC DR EMBL; U41849; AAB68254.1; -
CC DR SCD; S0006006; SEC16.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094 LYS-RICH.
FT MUTAGEN 1059 1059 L->S: IN SEC16-4; TS ACCUMULATION OF ER
FT MUTAGEN 1084 1084 L->P: IN SEC16-3; TS ACCUMULATION OF ER
FT MUTAGEN 1089 1089 L->P: IN SEC16-2; TS ACCUMULATION OF ER
FT MUTAGEN 1231 1231 W->R: IN SEC16-1; TS ACCUMULATION OF ER
FT CONFLICT 522 522 MISSING (IN REF. 1).
FT CONFLICT 560 560 I -> F (IN REF. 1).
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;

Query Match 3.5%; Score 185; DB 1; Length 2195;
Best Local Similarity 21.6%; Pred. No. 0.57;
Matches 223; Conservative 124; Mismatches 418; Indels 266; Gaps 54;

QY 72 GKVPYDALFSEELMKDPNQKADIVNEVKGK-----VLIKVDGK---- 113
DB 1283 GNPV-----LSNEP--VMADSVIFESIGNMNFISILWDEIYEIFSYDPKFGF 1331
QY 114 -----YVYIKDAHADNVTK--DEINRQ-KQEHVKDNEKN-----SNVAVARQS 157
DB 1332 SSILPKIYHASLQGLNSLGTQKTVTLSSSVRLKPKKDILTLNLTRELSVASRLSE 1391
QY 158 GRVTNDGVNENPA----DIIDTGNAYIVPHCG-----HHYIIPKSDL 197
DB 1392 ----SNTGLAKFLKSVGQDKSFNKYI---GGDDIDALNKKNDKKVFGFTPGCSA 1444
QY 198 SASLAAAKAHLAKGNQPSQLSYSTASDNNTQSVAKGST--SKPANKSENLQSLLEL 255
DB 1445 NSSTVDLTOTFTFQAQVTSQSVDVTTALLHNAHNVPSHVLHSPKFSNV-----KGL 1497
QY 256 YDPSAQVSESGLVFDPAKILSRTPNGVAIPH-----GDHYHPIPKSLALSEE 306
DB 1498 VEANLPTTHRIGDSLOGSPQR-IHNTQFAAAEPQMASLRVRVTDQHTNEKALKSQQILEK 1556
QY 307 KIARMPISGTSTVSTNAKPNVSWSLGSLSSNPSSLTSKELSSASDGYIFNPKDIYE 366
DB 1557 KSTAYTPQGNHSPVMEKSNVNSVLFADFAPP-----KLGIVPSNYVSSP-DLVR 1608
QY 367 ETATAVIVRHGDHFHYIPKSNQIGQPTLPNNS-----LATPS-PSLPINPCTS--HEKHEE 419
DB 1609 RES---IISTGSEFLPPPK---IGVPTKANSSQGLMSVSPSEALPIDVPVQVHETGYN 1662
QY 420 DGYGFANRIIADDES--GFVWSHGDNHNYFFKKDLTEQIKAAQKHLBEVTKSHNGLDS 477
DB 1663 DFNKKSQKSMPEDESHSHNSADQNTLKSDADVTDMDIEGPGFNDVK---NLLPM 1719
QY 478 LSHSEODYPGNAXEMKDKKIEKTAGIMQYGVKRESIVVKNKNAIYPHGHHDHAD 537
DB 1720 EPNHQPTSTVNPQI-----TISDDIQIL-QTNVEVRCTDASKMENS--PSIENERS- 1769
QY 538 PIDEHPVGIHSHSNYELFKPEEGVAKCKEKNVYTGEBLTVNVLNLLKNSTNNQNFLLA 597
DB 1770 --SEQFENISKASS--AYLSTGSLSLSE-NRPLTQDE--NSISETGVSTY-----LP 1816
QY 598 NGOKRVSFPPPELEKLGINMLVKLITPDGKVLKSVKGVGEGVGNTA----- 647
DB 1817 AG-----SISMEAKPISQVQDVPVNVNKNKSLYEQHMAKPKPKSTDAT 1860
QY 648 --NFELDQYLPQGTFTKYIIASKDPEVSDGTFTVPTSLAYKMASQ-----TIFYPF- 698

DB 1861 KNY---SPYVP-----QSTAASADGDE-----STILKTSPIAIRTHQAHASNPQYFFLV 1909
QY 699 -HAGDTYLRVNPQFAVPKGTDLVRVDFHGNAVLENNYKVGKELPIPKLN----- 750
DB 1910 NQANET-----ASFELSESTQA-----QSNNGVASENRFS-----PIKKAEEVVEKDTF 1953
QY 751 QGTRTAGNK-----LPVTFMANAYL-----DNSTYIVVEVPILEKENQTKPSIL 796
DB 1954 QPTIRKASTNQYRAPKPLESDADKYNDVLEDDSDNNSTDEAKNRKEEKNVNNKKETK 2013
QY 797 POFKRKKAQENSKLDEK-----VEEPTKSEKVEKEKLSGTSNSTSTLEEVPTVPV 849
DB 2014 P-----SNKDIDDKSNGWEGWLLKKTGDKVYKAKLGHKNTLIYDEKLRWYKDAT 2065
QY 850 QEKVAKPAESYGMKLENVLFNMD-----GTIELYLP-----SGEVIKKNMADTGE-- 895
DB 2066 EEEKOKIIESSAPPPPIVKKRKGDKGPKTPRSGPINNSLPPVHATSVIPNN--PITGEP 2123
QY 896 -----APOG-NGENKPSNGKVS--TGTVENQPTENKPADSLPEAPNEKP--VKPENST 944
DB 2124 PIKTSPTGTPGNPNNSPSPSSPISISGVNLTSKRANGDLLLSLGGPKPASTRRKKKT 2183
QY 945 DNGMLNPEGNV 955
DB 2184 ARGVNVNMDNI 2194
RESULT 11
SPAZ_YEAST
ID SPAZ_YEAST STANDARD; PRT; 1466 AA.
AC P23201;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SPAZ protein.
GN SPA2 OR PEAL OR YLL021W OR LI209.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE-91009481; PubMed-2211820;
RA Gehring S., Snyder M.;
RT "The SPA2 gene of Saccharomyces cerevisiae is important for
RT pheromone-induced morphogenesis and efficient mating."
RL J. Cell Biol. 111:1451-1464(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Purnelle B., Goffeau A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
CC EFFICIENT MATING, PERHAPS AS A CYTOSKELETAL PROTEIN.
CC -!- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMOO TIP
CC (MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
CC -!- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X53731; CA37763.1; -
DR EMBL; X97560; CA66170.1; -
DR EMBL; Z73126; CA97469.1; -
DR PIR; A36426; A36426.
DR SGD; S0003944; SPA2.

Biochim. Biophys. Acta 1075:248-258(1991).

- CONNECTION: May play a role in intercellular signaling and in connecting cells with the extracellular matrix. May take part in the regulation of cell motility, growth and differentiation. Binds hyaluronic acid.

- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2 and V3; are produced by alternative splicing.

- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed in the central nervous system, and in a number of mesenchymal and epithelial tissues; the major isoform V2 is restricted to the central nervous system.

- DEVELOPMENTAL STAGE: Disappears after the cartilage development (By similarity).

- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

- SIMILARITY: CONTAINS 2 LINK DOMAINS.

- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF060456; AAC24358.1; -
EMBL; AF060457; AAC24359.1; -
EMBL; AF060458; AAC24360.1; -
EMBL; AF060459; AAC24361.1; -
HSSP; P01132; IEPG.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF000008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR Pfam: PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 1.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR GlycoProtein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN CORE PROTEIN.
FT DOMAIN 37 138 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 168 245 LINK 1.
FT DOMAIN 266 347 LINK 2.
FT DOMAIN 266 347 LINK 2.

FT	DOMAIN	349	1336	GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
FT	DOMAIN	1337	3074	GAG-BETA.
FT	DOMAIN	3074	3110	EGF-LIKE 1.
FT	DOMAIN	3112	3148	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	3161	3275	C-TYPE LECTIN.
FT	DOMAIN	3280	3338	SUSHI.
FT	DISULFID	44	131	BY SIMILARITY.
FT	DISULFID	173	244	BY SIMILARITY.
FT	DISULFID	197	218	BY SIMILARITY.
FT	DISULFID	271	346	BY SIMILARITY.
FT	DISULFID	295	316	BY SIMILARITY.
FT	DISULFID	3078	3089	BY SIMILARITY.
FT	DISULFID	3083	3098	BY SIMILARITY.
FT	DISULFID	3100	3109	BY SIMILARITY.
FT	DISULFID	3116	3127	BY SIMILARITY.
FT	DISULFID	3121	3136	BY SIMILARITY.
FT	DISULFID	3138	3147	BY SIMILARITY.
FT	DISULFID	3154	3165	BY SIMILARITY.
FT	DISULFID	3182	3274	BY SIMILARITY.
FT	DISULFID	3250	3266	BY SIMILARITY.
FT	DISULFID	3281	3324	BY SIMILARITY.
FT	DISULFID	3310	3337	BY SIMILARITY.
FT	CARBOHYD	57	57	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	817	817	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	965	965	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1017	1017	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1333	1333	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1393	1393	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1463	1463	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1653	1653	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	1974	1974	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2045	2045	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2074	2074	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2103	2103	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2263	2263	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2356	2356	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2623	2623	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2641	2641	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	2919	2919	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	3354	3354	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	3364	3364	N-LINKED (GLCNAC.) (POTENTIAL).
FT	CARBOHYD	349	349	P -> R (IN ISOFORM V1 AND V3).
FT	VARSPLIC	349	349	MISSING (IN ISOFORM V1).
FT	VARSPLIC	350	1336	MISSING (IN ISOFORM V2).
FT	VARSPLIC	1337	3074	MISSING (IN ISOFORM V3).
FT	VARSPLIC	350	3074	MISSING (IN REF. 2).
FT	CONFLICT	25	25	MISSING (IN REF. 2).
FT	CONFLICT	51	51	MISSING (IN REF. 2).
FT	CONFLICT	89	89	N -> D (IN REF. 2).
FT	CONFLICT	96	96	Q -> D (IN REF. 2).
FT	CONFLICT	346	346	C -> R (IN REF. 2).
FT	CONFLICT	3381	3381	AA; 369984 MW; F09716PFA778D459 CRC64

Query Match	3.5%	Score 184;	DB 1;	Length 3381;
Best Local Similarity	20.1%	Pred. NO.	1.2;	

```

Qy 233 VAKGSTKPAKNSENLQSLKE-LYDPSAQRYSESGL-VEDPAKIISRTNGVAIPHG 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1649 L$ORNTIDPRVTSDT$KIMITSLVDVPTTIYSIQ$Q$AVVTKFVREI----- 1699
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 291 DHYHFI--PY$KLSALE$K$IARMVPI$GT$TV$TNA-----KPN$EV$SS$G$LS 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1700 DTY$EW$F$PELE$TRKE$EKG--TTGTASTVEH$PTQRLDQ$V$P$PELESS--$ 1751
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 339 SNP---$SLT$K$EL$S$AS$G$FI$PKD$IVE$TAT$YIVR$GD$HF$YIP$K$N$QIGQ$TLP 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1752 ET$P$D$D$A$A$TR$K$FT$Q-----MPTQ$ER$T$T$ST$V$F-----K$E$V$--LD 1793
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 396 NNSLATPSPSL-----PINPGTSHKHEEDGYGFDA-----NRI 429
DB 1794 NLAAQTDPDLSLSSQGVLEVSTVPVSGVSLFMEQSGEAAVDPETTVTSLSLNIEPEI 1853
QY 430 IADESSGVFNSHGDNHVFYFKKDLTEQIQKAAKHLEEVKTSNGLDLSLSEHQDYPGNA 489
DB 1854 LAKEBAAGASPNVETVPPFP--TEQVLSTAVDREVAETISQTSKNLVSEISGEPTHR 1911
QY 490 KEMK--DLDKKIEEKIAGIKQYGVKRESIVYVKNKNAIIPYHGDHHDH--537
DB 1912 AEIKGFSTDFPLEEDFSDGFREYST--VSPYITKEEIVVMWEGSGDAAFKDTOMLPSVTPT 1969
QY 538 -----PIDEKHPGVIGHSHSNLYELFKPEGVAKKEGKKNVYTGELTNVNLKNKTFNN 591
DB 1970 SLSLNTADSSEP---GSTLVSTAFWEPTASAEG---SGEPLLSV-----SSVD 2016
QY 592 QNFTLANGKRVSFPPPEKKLG-----INML-----VKLITPDGKYLE- 632
DB 2017 QVFPAAAG--KASGTDSPFIDQRLGEGAINETDQRTSTILPTAEASTKASTEEGEVKEN 2074
QY 633 -----KVSCKVTGEGVGNIA-----NFE-LDQYPLPGQTFKY 663
DB 2075 HTVSMDFPTEVPDELWPRQVNPVROGNGSEIVSEKTEQESFEPLOSSVAPEQT--- 2131
QY 664 TIASKDYPE--VSYDGTFTVPTSLAYX-----688
DB 2132 TFDSTQTFPEGLQTTGYFTLTTRKTYSTDERMEDEVISLADVSTPTLDSKGLVLYTTLPE 2191
QY 689 -----MASOTIFVPRH---AGDTYLRVNPQFAPVKGTDALVRVDE---EHNAY 732
DB 2192 VTEKSHFLATASVETSPAESVIAGSTIKESIKFPFKVTPSIKESDTOLIFSLGSG 2251
QY 733 LE-----NNTKVGGEIKLPIKLNQGTTR-----TACNKIPVTPMAN----- 768
DB 2252 GEEVLTLGVSNFTEIQVLSLTPLTSQVOSLEASILNDTSGDYEGMENVANEMRPLIS 2311
QY 769 -----AYLDNQSTYIIVEPILSEKNOTDKPSILP--OPKRNKAQENSKLDEKVEEPTKSEV 823
DB 2312 KDSIFDEGETASSTTLEILSDARTGPTAPLTFGTGQPNQONTHRAEBIQIS--- 2368
QY 824 EKEKUSGTNSTSLTEVTPDPOVEKVAFAESYGMKL-----ENVLFNMD 872
DB 2369 RPQPLTDQVSSNSVTATKETAATPATDFLAR---TYDLEMAKGFVTPTPKPSDLFEVHS 2425
QY 873 GTIELYLSGVIKKNMADTFGEAPOGNGENKPSKENGKVTGTVENOPTENKPADSLPEA 932
DB 2426 GE-----CSGEL-----DAGAEVHASGMTQATROG--STTFVSDRSLEKH-----PKV 2467
QY 933 PNEKQVKNPENTDNGM-----LNPEGNGVS-----DMLDPALAEAP--AVDPQVE- 976
DB 2468 PSVEAV-----TVNGFPTVMVLPRLHPEQREGSPTEATGTPASTASYEKATEGAADSFQDH 2522
QY 977 -----KLEKFTASYGLGLDSVIFNMD 997
DB 2523 FWGFKDSTLKPDKRKAT-----ESILIDLD 2547

RESULT 13
ID MSPL_PLAF3
AC P19598; 025921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMSA) (p190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.

```

```

RX MEDLINE-88166657; PubMed-3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RL precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RN EMBO J. 6:4137-4142(1987).
RP SEQUENCE OF 1032-1582 FROM N.A.
RA MEDLINE-95354793; PubMed-7628566;
RT Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RL merozoite surface antigen-1.";
RC Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35727; AAA29715.1; -
CC EMBL; Y00087; CAA69280.1; -
CC EMBL; Z35326; CAA84555.1; -
CC PIR; S06286; S06286.
CC InterPro: IPR000561; BGF-like.
CC Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82AJE159948CAD6 CRC64;

Query Match 3.4%; Score 181.5; DB 1; Length 1682;
Best Local Similarity 20.1%; Pred. No. 0.57;
Matches 238; Conservative 173; Mismatches 417; Indels 359; Gaps 63;

QY 3 YALNQRSQEN----KNNRVSIVYDGSQSKSENITPDQVSKQKGIQAEQIVKTIIDOG 58
DB 511 YTVNVEKQRYNNKFSSNNVYVNV---QKLKALSYLED-YSLRKG-----553
QY 59 YVTSHGDRHYHYNGKVEYDA---LFSBEL-----LMKDPNYQLKADIVN- 100
DB 554 ---SEKDFNHYITLKTGLEADIKKLTGEEKSENKILEKNKGLTHSANASLEVSIVKL 610
QY 101 EVKGGYII-KVDG--KYVYVYLKAAHADNVNT-----KDEINRQKQ- 138
DB 611 QVQVLLIKKIEDLRKLEFLKNAQLKDSHVNPNIYKPNKPPYVILVKREVDKLEF 670
QY 139 ----EHWKNEKVNNSVA---VARSGRYTTNDGVVFNPAIDIEDTGNAYIVPFGGHHY 190
DB 671 IPKVYKMLKKEQAVLSITQPLVAASE---TTEDG-----GHSTHTLSQSGETE 716
QY 191 YIPKSDLSASELAATAKHAHLAKGNQWPSQSLSYSTASDNNTQSVAKGSTSKPANK----- 244
DB 717 VTEETEETVGHVTTTTLTPPKVEKVVNSIEHKND--NSQALTQVYLKLDLDELFTLSY 775

```



```

Db 969 PQPVTENTSNEKNVP-----KADQSTKQDNISNFADVEVSASSYPKLDKDAETDDQILIG 1024
Qy 460 AAKQ-----HLEVKTSNGLDLSLSSHEQDPCNAKEMKDLID-----KKI 499
Db 1025 AINSSSVHGKELPDLAETVTTA-NRVSTTSAGDQIDTQSEFLRAADVENLSDDDDEHRQN 1083
Qy 500 EKIAG-----IMKQGVVRESIVNKNKNAIYPHGDDHHADPIDEKHPVIGCHSHS 552
Db 1084 ESRVNDOLFQDIPKHYRNQENI--NSSN--LFPEG-----KVPQEBK--GVSNENI 1131
Qy 553 NYELFKPEGVAKKGNKVYTGELTINVLNKKNSTNNQ-----NETLANGOK 601
Db 1132 NISLKTNEDATLQKLSPOASKVLTENSNELKDT--NNSGKDAKDIKLDDYSDKETAK 1189
Qy 602 RVSFEPPE-----LEKKGILNMLVLTIPDGKVLKSVGKVGEGVGNIANFELDQPY 555
Db 1190 EIT--KPKNFVEGITERKEIFPTIPRLAPPASKI-----NFQSPSY 1229
Qy 656 LPOQFKYTIASKDYPEVSYDGTFTVPTSLAYKNASOTIFYPFHAGDTYLRVNPQFAVPK 715
Db 1230 I--ELFQGRVRLDKHDAHYNK-----RLASQVSL-----SEGKVN----- 1265
Qy 716 GTDALRVFDEFGHNAVLENNYKGEI-----KLPIPKLNOGTRTGNKIPVTWNAVY 770
Db 1266 --FEEDAATINKSQDAAKARTQISEVIEYEMQOPIP-----TYLPKAK 1308
Qy 771 LDNQTYIIVEPILEKENQTKPSLPQFRNKAQENSKLDERVEPKTSEKVEKELSE 830
Db 1309 LDDSG-----LEKSD-----DKFFEIEELKEELKSGKTGDNEDVGN 1344
Qy 831 TGNSTNSTLEVP-----VDPQVQKVAFAESYGMKLEN-----VLFNMD----- 872
Db 1345 NNPSSIBKIKPKPAKVFIRTPSVRIIGRUFEDT--RKYENGSPDISFTYDFTHNDEPD 1402
Qy 873 -GTIELYPSGEVKKNNMADFTGEAPQNGENKPSNGKVTSTGVNQENPKNPADSLPE 931
Db 1403 KRLMELKPSQDEIPDDR--YT-----PAE-----EPTAEFPVEELPN 1439
Qy 932 ANEXPKVPEN--STDNGMLNPEGVNSDP--MLD 962
Db 1440 TPRSINVTTSNKSSTDDKL--SSGNIDQKTELLD 1472

RESULT 15
PGCV_RAT
ID PGCV_RAT STANDARD: PRT: 2738 AA.
AC Q9EBR4: 008592; 088564; Q9RLK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vesican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (pg-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CP5G2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=93327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wright T.N.;
RT "Versican/pg-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=93308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,

```

```

RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RL brevican.";
RN Biochem. Biophys. Res. Commun. 247:207-212(1998).
RC [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE=Kidney;
RX MEDLINE=96094159; PubMed=9434070;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney.";
RN Nephron 77:461-470(1997).
RC [4]
RP SEQUENCE OF 2535-2738 FROM N.A.
RC STRAIN=Soraguet-Dawley; TISSUE=Lung;
RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
RT "Molecular cloning and characterization of two developmentally
RT regulated genes in rat lung.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
CC Vint; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
CC but not in glomeruli.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF062402; AAC40166.1; -
CC EMBL: U75306; AAB51125.1; -
CC EMBL: AF084544; AAD48544.1; -
CC EMBL: AF072892; AAC26116.1; -
CC EMBL: AY007891; AAG16831.1; -
CC HSSP: P01132; IEPG.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF_2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003600; Ig-like.
CC InterPro: IPR001304; Lentin_C.
CC InterPro: IPR000538; Link.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC PRINTS: P01265; LINKMODULE.
CC SMART: SM00032; CCP; 1.
CC SMART: SM00034; CLECT; 1.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00179; EGF_CA; 2.
CC SMART: SM00409; IG; 1.
CC SMART: SM00410; IG_LIKE; 1.
CC SMART: SM00445; LINK; 2.
CC PROSITE: PS00010; ASX_HYDROXYL; 1.
CC PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
CC PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
CC PROSITE: PS00022; EGF_1; 2.
CC PROSITE: PS01186; EGF_2; 1.

```

```

DR PROSITE: PS01187; EGF CA: 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 2738
FT FT NON_CONS 348 349
FT FT DOMAIN 37 137
FT FT DOMAIN 167 244
FT FT DOMAIN 265 346
FT FT DOMAIN <349 695
FT FT
FT DOMAIN 696 2431
FT FT DOMAIN 2431 2467
FT FT DOMAIN 2469 2505
FT FT DOMAIN 2518 2632
FT FT DOMAIN 2637 2695
FT FT
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 345
FT DISULFID 294 315
FT DISULFID 2435 2446
FT DISULFID 2440 2455
FT DISULFID 2457 2466
FT DISULFID 2473 2484
FT DISULFID 2478 2493
FT DISULFID 2495 2504
FT DISULFID 2511 2522
FT DISULFID 2539 2631
FT DISULFID 2607 2623
FT DISULFID 2638 2681
FT DISULFID 2667 2694
FT FT
FT CARBOHYD 57 57
FT CARBOHYD 330 330
FT CARBOHYD 692 692
FT CARBOHYD 758 758
FT CARBOHYD 802 802
FT CARBOHYD 805 805
FT CARBOHYD 1257 1257
FT CARBOHYD 1435 1435
FT CARBOHYD 1633 1633
FT CARBOHYD 1660 1660
FT CARBOHYD 1684 1684
FT CARBOHYD 1738 1738
FT CARBOHYD 1848 1848
FT CARBOHYD 2004 2004
FT CARBOHYD 2409 2409
FT CARBOHYD 2711 2711
FT CARBOHYD 2721 2721
FT FT VARSPLIC 349 2431
FT FT VARSPLIC 2697 2738
FT FT
FT CONFLICT 2535 2539
FT SEQUENCE 2738 AA; 300004 MW; 12CA626D58BD8C6A CRC64;
Query Match 3.4%; Score 178.5; DB 1; Length 2738;
Best Local Similarity 19.5%; Pred. No. 1.6;
Matches 223; Conservative 152; Mismatches 391; Indels 375; Gaps 58;
QY 19 VSY-VDGSSQSKSNLTPDQVSOKEGIAQIVTKITDQGVVTHSGDHYHYNGKYPYD 77
DB 501 VSYAEGSSPWEFASTLP-----FGQVQLVEETEG-----KKPLD 541
QY 78 -----ALFSELLMKDNYOLKADIVNEVKGYYIKVDGKYYVYLKDAHADNVRTKD 131
DB 542 YTDLSGLFEQPRVTELPDFSMTPSDI-----SVFTADLSLHRTPLRPPS 587
QY 132 EINROKQEHVD-NEKVNNAVAVR---SQGRYTTNDGYVNPADIETDGTNAYIVPHG 187
DB 588 PTEEPHIFEKPESEKTTGDIILPRESVTOHPLT-----LMDIITAKKTESDI-- 637

```

```

QY 188 HYHYIPKSDLSASELAAGAHLAGNMOPSQLSYSTASDNNTQSVAKGSTSKP----- 241
DB 638 EVHMTSKDPV-----MQPTRSVVVERTTSKPQL-----SISPPAGTKF 678
QY 242 -----ANKSENLOSLKELYDSPSAQRYSESDGLVDP-----AKISRTPNG 284
DB 679 HPDINVIIEVRENKTGRMSDMVNGHPIDSEKKEEPCSEETDPLHDLFAELPELPS 738
QY 285 VAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNWSSLSGLSNPSSL 344
DB 739 FEI---DIYH---SEDEDEGEDCVDNATDVTTPSVQITCKPH-----VTVKPNPEA 787
QY 345 TTSKEL-----SSASDGVIFNPKDI-----VEETAYIVRHGDHFH 381
DB 788 EARRGLYESVAPSNFNTSATDTHQFIPAETELSTMQFTKSKATELLEITWKPTYP 847
QY 382 YIPKSNQIQG-----PTLPNNSLATPSPLPI---NPGVSHKHEHD----- 420
DB 848 ETPHFSSGEPDVPFTLPDHGDKTKWKSEFTESNPNTENPEHKQPKPIPLPEFSSEG 907
QY 421 -----GYGFANRIAD--ESGVMSHGDNHNYFFKKDLTEEQI- 458
DB 908 AIDQASQOTIFSRATEVALGKETDQSPITSTSSIRSGSVSHA-----LEEDPIA 957
QY 459 -KAACKHLEEVKTSIHNGLDLSLSSHEQDYPGNK-----EMKDLKKIEKTAGINKQ 509
DB 958 LTGTSQTDSESTVESVEMTPSQTFESGSSAPTIEGSGEVEEYTNKIFNTVDLPQR 1017
QY 510 YGVKRESIVNKEKNATIIYPHGDH-----HADPID-----EHKPVGIG----- 550
DB 1018 --EPTDTIPLDMSNIMI---TDHHIYTPATTAPLDSOLPSTDARPTQFGIQTTSWVS 1072
QY 551 HSNVELFPPEGVAKKEGNKYVTGS-----ELTVNVLNKKSTNNQNFLLANGKRVSF 605
DB 1073 STSPEGRTEED-KERDTNAAHTGEVOPATERSD--RLLTSELESSN-----VAA 1120
QY 606 SFPPELEKKGGINMLVLIITPDGKVLKVS-----GKVFEGVGNTANFELDQPY 655
DB 1121 SSP-----LDTWEGFVPEITTSVSEKEMANTTPVTE-TSDVANLET----- 1161
QY 656 LPQTFKTIASKDYPEVSYDGTFTVPTSLEYKMASOTIFYFHAGDTYLRVNPQFAVPK 715
DB 1162 ---OSFESSSSQ--PRVQEE-----LTTLSGK--PPLIFMDLGGDA----- 1197
QY 716 GTDALVRVDFEHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPVTPMANAYLDNQ 775
DB 1198 STDMEFITASSF--TLDESQTKV---KKELPSTLSFVSSTSSSEFGLAPSTVLD--- 1249
QY 776 TYIIVEPILEKENOTDKPSILLPQFKRNKAQENSKLD-----EKVEEPKTSKVE 824
DB 1250 -----IEIVEVMNOTSKRLTISELSGKPTSQAEVRDLYPGLGEDFSGDSSEYPTVSSTTM 1304
QY 825 KEKLSETGNSFS-----NSTLEEVP---TVDPVQ-----EKVAKEAE 858
DB 1305 KEEIVGSGSENERVKDTQTLLSSIPPTSDNINPVDSKGFSTVASTATPAWFEFMTSAE 1364
QY 859 SYGKMLENLNFMMDGTIELXLPDSGEVIKKNMADFTGEAPQNGENKP-----SENGK----- 910
DB 1365 GSGBELSSV-----RSSVSLVPLGVLDILP-----TTSPYFDQEEFAAAVTEAGQSAL 1415
QY 911 ---YSTGTVENQPTENKPAD-----SLPEA---PNEKPVKPNSTNDGMLNPE-GNVGS 957
DB 1416 PIAYSGNTVD--LTENRDIENVSTMSVDLPQTMEPAKLWSKPE-----VNPKEQIGS 1466
QY 958 D 958
DB 1467 E 1467

```

Search completed: May 13, 2003, 13:54:54
Job time : 32.9531 secs

Blank body text area

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 45.2509 Seconds
(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAVALNQHRSQENKNNRVS.....TELRLPSGEVKKNLSDPIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	5297	99.7	1039	16 Q9ANY1	Q9any1 streptococc
2	1279.5	24.1	844	2 Q9AG74	Q9ag74 streptococc
3	1256	23.6	816	2 Q9AHT9	Q9aht9 streptococc
4	1255	23.6	802	16 Q97QM8	Q97qm8 streptococc
5	1235.5	23.2	819	16 Q97QM9	Q97qm9 streptococc
6	1233	23.2	839	16 Q9ANY2	Q9any2 streptococc
7	1232.5	23.2	819	2 Q9ANY3	Q9any3 streptococc
8	902	17.0	825	16 Q99XV4	Q99xv4 streptococc
9	899.5	16.9	822	2 Q9ZHG7	Q9zhg7 streptococc
10	897	16.9	825	2 Q93GT5	Q93gt5 streptococc
11	661.5	12.4	289	2 Q9AE21	Q9ae21 streptococc
12	292.5	5.5	792	16 Q99Z76	Q99z76 streptococc
13	238.5	4.5	2004	16 Q97QP7	Q97qp7 streptococc
14	235.5	4.4	1078	5 Q963T1	Q963t1 plasmodium
15	228	4.3	2485	5 Q96134	Q96134 plasmodium
16	226.5	4.3	1139	5 Q97237	Q97237 plasmodium

17	223	4.2	1964	2 Q59947	Q59947 streptococc
18	218	4.1	1873	2 Q924N7	Q924n7 enterococcu
19	216	4.1	5458	5 Q90459	Q90459 plasmodium
20	214	4.0	1881	16 Q8RGK2	Q8rgk2 fusobacteri
21	214	4.0	1939	5 Q25662	Q25662 plasmodium
22	212.5	4.0	1733	2 Q9K114	Q9k114 staphylococ
23	210	4.0	1694	5 Q9NHX1	Q9nhx1 plasmodium
24	209	3.9	1276	2 Q93TY6	Q93ty6 staphylococ
25	207.5	3.9	1694	5 Q9TZT5	Q9ztz5 plasmodium
26	206	3.9	1593	5 Q8SX82	Q8sx82 grosophila
27	204	3.8	1368	16 Q93M90	Q93m90 clostridium
28	204	3.8	1524	10 Q8RYN2	Q8ryn2 oryza sativ
29	204	3.8	2647	5 Q9U4X0	Q9u4x0 plasmodium
30	203.5	3.8	1704	5 Q9TZT4	Q9ztzt4 plasmodium
31	199	3.7	1720	5 Q25922	Q25922 plasmodium
32	198	3.7	1179	2 Q9AHK5	Q9ahk5 borrelia bu
33	196.5	3.7	881	2 Q8RFV4	Q8rfj4 streptococc
34	196	3.7	1271	5 Q25860	Q25860 plasmodium
35	195.5	3.7	1927	2 Q54875	Q54875 streptococc
36	195.5	3.7	2910	10 Q9FND5	Q9fnd5 arabidopsis
37	194.5	3.7	2806	16 Q8RI19	Q8ri19 fusobacteri
38	193	3.6	3724	5 Q77320	Q77320 plasmodium
39	192	3.6	5005	16 Q9PP25	Q9pp25 ureaplasma
40	190.5	3.6	1395	2 Q9AISO	Q9aiso staphylococ
41	190.5	3.6	4667	5 Q9TVI9	Q9tvi9 caenorhabdi
42	190	3.6	1698	2 Q9LC00	Q9lc00 staphylococ
43	189.5	3.6	3257	5 Q9V736	Q9v736 grosophila
44	189	3.6	1038	16 Q99RD2	Q99rd2 staphylococ
45	189	3.6	1236	5 Q9GTX2	Q9gtx2 plasmodium

ALIGNMENTS

RESULT 1

Q9ANY1	PRELIMINARY;	PRT; 1039 AA.
AC Q9ANY1:		
DT 01-JUN-2001 (TRENBLrel. 17, Created)		
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)		
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)		
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).		
DE SP1004.		
GN PHE OR SP1004.		
OS Streptococcus pneumoniae.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group: Lactobacillales;		
OC Streptococcaceae; Streptococcus.		
OX NCBI_TaxID=1313;		
RN [1]		
RP MEDLINE=21101045; PubMed=11159990;		
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,		
RA Dormitzer M., Degan R., Brewah Y.A., Barren P., Lathigra R.,		
RA Langermann S., Koenig S., Johnson S.;		
RT "Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis.,"		
RL Infect. Immun. 69:949-958(2001).		
RN [2]		
RP SEQUENCE FROM N.A.		
RA MEDLINE=21357209; PubMed=11463916;		
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA Durkin A.S., Whittam M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,		
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,		
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;		
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.,"		
RL Science 293:498-506(2001)		
DR EMBL; AF318956; AAR06761.1; -		

DR EMBL: AE007403; AAK75121.1; --
 DR TIGR: SP1004; --
 KW Signal: Hypothetical protein; Complete proteome.
 FT SIGNAL 1 29 POTENTIAL.
 SQ SEQUENCE 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match 99.7%: Score 5297; DB 16; Length 1039;
 Best Local Similarity 99.7%: Pred. No. 1.1e-241;
 Matches 1016; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAEQIVIKITDOGVY 60
 DB 21 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAEQIVIKITDOGVY 80

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 120
 DB 81 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 140

QY 121 AAHADNVRTKDEINRQKQEHYKDKNEKYNVAVARSQGRYTTNDGVYFNPADIIEDTGA 180
 DB 141 AAHADNVRTKDEINRQKQEHYKDKNEKYNVAVARSQGRYTTNDGVYFNPADIIEDTGA 200

QY 181 YTVPGHGHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDNNTQSVAKGSTK 240
 DB 201 YTVPGHGHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDNNTQSVAKGSTK 260

QY 241 PANKSENQSLKELYDPSQRYSESDGLVDPAKIISRTNGVAIPGHGDHYHFYPSK 300
 DB 261 PANKSENQSLKELYDPSQRYSESDGLVDPAKIISRTNGVAIPGHGDHYHFYPSK 320

QY 301 LSALAEKTAARVPIISGTGSTYSTNAKNEVSVSLGSSNPSSLTTSKELSSASDGYIFN 360
 DB 321 LSALAEKTAARVPIISGTGSTYSTNAKNEVSVSLGSSNPSSLTTSKELSSASDGYIFN 380

QY 361 PKDIVEETATAYIVRHGDHYHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHKHEED 420
 DB 381 PKDIVEETATAYIVRHGDHYHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHKHEED 440

QY 421 GYGFDANRIAEDESGFVMSGHDHNYFFKDLTEEOIKAAQKHLSEVKTSHNGLDSLSS 480
 DB 441 GYGFDANRIAEDESGFVMSGHDHNYFFKDLTEEOIKAAQKHLSEVKTSHNGLDSLSS 500

QY 481 HQDTPGNKAKEMKDLKIEKIAKIQKQYKVPRESIVWKNENAIYIPGHGDHHDADPID 540
 DB 501 HQDTPGNKAKEMKDLKIEKIAKIQKQYKVPRESIVWKNENAIYIPGHGDHHDADPID 560

QY 541 EHKPVGIGHSHSNVELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNFILANGQ 600
 DB 561 EHKPVGIGHSHSNVELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNFILANGQ 620

QY 601 KRVSFSPPELEKLGINMLVLTIPDGKVLKGVKGVGEGVGNIANFELDOPYLPQGT 660
 DB 621 KRVSFSPPELEKLGINMLVLTIPDGKVLKGVKGVGEGVGNIANFELDOPYLPQGT 680

QY 661 FKYTIASXDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFVAPKGTDAL 720
 DB 681 FKYTIASXDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFVAPKGTDAL 740

QY 721 VRVDFEFGNAYLENNYKVGIEKILPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIYE 780
 DB 741 VRVDFEFGNAYLENNYKVGIEKILPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIYE 800

QY 781 VPILKENOTDKPSILPOFKRKAQENSLKDEKVEPKTSEKVEKELSGTNSNSTL 840
 DB 801 VPILKENOTDKPSILPOFKRKAQENSLKDEKVEPKTSEKVEKELSGTNSNSTL 860

QY 841 BEVPTVDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNADTFGEAPQGN 900
 DB 861 BEVPTVDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNADTFGEAPQGN 920

QY 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 960
 DB 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 980

QY 961 LDPALAEAPAVDPVOEKLKFTASYGLGLOSVIFNMDGTIELRLPSEGEVYIKKNLSDFIA 1019
 DB 981 LDPALAEAPAVDPVOEKLKFTASYGLGLOSVIFNMDGTIELRLPSEGEVYIKKNLSOLIA 1039

RESULT 2

Q9AG74 PRELIMINARY; PRT: 844 AA.

AC Q9AG74; AC Q9AG74; (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PhpA.
 GN PhpA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 CC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21246685; PubMed=11349048;
 RA Zhang Y., Masi A.W., Barniak V., Mountzouros K., Hostetter M.K.,
 Green B.A.;
 RT "Recombinant PhpA Protein, a Unique Histidine Motif-Containing Protein
 from Streptococcus pneumoniae, Protects Mice against Intranasal
 Pneumococcal Challenge.";
 RL Infect. Immun. 69:3827-3836(2001).
 DR EMBL: AF340221; AAK26629.1; -;
 SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;

Query Match 24.1%: Score 1279.5; DB 2; Length 844;
 Best Local Similarity 40.5%: Pred. No. 2e-52;
 Matches 301; Conservative 101; Mismatches 182; Indels 159; Gaps 21;

QY 1 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAEQIVIKITDOGVY 60
 DB 20 CSYELGLYKQVATKNNRVSYIDGKQATQNTLTPEVSKREGINAEQIVIKITDOGVY 79

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 120
 DB 80 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 139

QY 121 AAHADNVRTKDEINRQKQEHYKDKNEKYNVAVARSQGRYTTNDGVYFNPADIIED 176
 DB 140 AAHADNVRTKDEINRQKQEHYKDKNEKYNVAVARSQGRYTTNDGVYFNPADIIED 199

QY 177 TGNAYIVPHGHHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDN----- 228
 DB 200 TGDAYIVPHGHHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDN----- 259

QY 229 -----NTQSVAKGSTKPAKSENLSQSLKELXDPSQAQRYSESDGLVDPAKIISR 280
 DB 260 SVSNPGTNTNTNSNTNSQASQSDNDISLLQYKLPQSRHVESDGLIFDPAQITSR 319

QY 281 TPNGVAIPGHGDHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDN----- 328
 DB 320 TANGVAIPGHGDHYHYIPKSDLSASLAALAAKAGKKNMPSQSYSTASDN----- 379

QY 329 --EVVSLGSLSSNPSSLTTSKE--LSSASQGYIFN-----PKDIVEETAT---AVI 373
 DB 380 SPQAPNPQAPSPNPIDEKLVKAVRGVGDGYVFEENGVPYIPAKDLSAETAAGDSKL 439

QY 374 VRHGDHYHY-----IPKSN-----QIGOPTLPNNSLATPSPSL----- 406
 DB 440 AKQESLHKLCAKKTDLPPSSDRFYKAYDOLLARIHQDLDLNDKGRQVDFEALDNLERLK 499

QY 407 -----PI-----NPGTSHEKHE-----EDGYGFDAN 427
 DB 500 DVSSDKVKLVDDILAFAPTRHPRERLGRKNAQTYTDDIEIQVAKLAGKYTTEDGYIFDPR 559

QY 428 RIIDEDESGFVMSGHDHNYFFKDLTEEOIKAAQKHLSEVKTSHNGLDSLSSHEQDYPG 487

Db 560 DITSEGDGAYVTPHMTSHWIKDLSSEARAAQAYAKE-----KGLTPPSTHQD-SG 613
QY 488 NAKEMKLDK-----KTEEKIAGIMKOYGVKRESIVVNKKNKNAIIPHGDDHHADPIDEH 542
Db 614 NT-SAKGAEAIYNKRAAKVPLDRMPYNLQ---YTVVEKNGSLIIPHYDHYH----- 662
QY 543 KPVGIGHSHSNYELFKPEEGVAKKGNKVTGDELTVNVLN-----KNSTENNON 593
Db 663 -----NFKFEW---DEGL--YEAPKGSLEDLATVYVVEHNERPHSDNGFCNAS 710
QY 594 FTL---ANGOKRVSFSPPELEK 613
Db 711 DHVORNNGOADTNQTEKPNEEK 733

RESULT 3
Q9AHT9
ID Q9AHT9 PRELIMINARY: PRT: 816 AA.
AC Q9AHT9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN pHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA;
RX MEDLINE=21116976; PubMed=11179332;
RA Wizenmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
Gayle A., Brewah Y.A., Walsh W., Warren P., Lathigra R., Hanson M.,
Langemann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1539-1598(2001).
DR EMBL; AF291695; AAK19155.1;
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 23.6%; Score 1256; DB 2; Length 816;
Best Local Similarity 41.3%; Pred. No. 2.4e-51;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQREGIOAEQIVIKITDQGYV 60
Db 20 CSYELGLYQARTVKNNRVSYIDGKQATKNTENLTPEVSKREGINAEQIVIKITDQGYV 79
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIYNEVKGGLIKVDGKYVYLKD 120
Db 80 TSHGDHYHYNGKVPYDAIISEELLMKDPNYKLKDEDIYNEVKGGLIKVDGKYVYLKD 139
QY 121 AAHADNVRTKDEINRQKQEHVKDNE----KVNSNVAVARSQGRYTTNDGYVFNPAIDIED 176
Db 140 AAHADNVRTKEELNRQKQEHVSHRGEGTFRNDGAVALARSQGRYTTDGYVFNPAIDIED 199
QY 177 TGNAYIVPGGHVHYTPKSDLSASELAALAAKHAHLAKNMQPSQLSYSSSTASDN----- 228
Db 200 TGDAYIVPGHDHYTPKSELASELAALAAAFISCRGNLSNRTYRRQNSDRTSRTNWVP 259
QY 229 -----NTQVAKGSTKPAKNSLNQSLKELYDSPSAORYSESGDLVFDPAKILSR 280
Db 260 SVSNPGTNTNTNSNTNSQASQNDIDSLKQLYKPLSORHVESGDLVFDPAQITSR 319
QY 281 TPNGVAIPGHDHYHFTPYSKLSALEKTIARMVP----- 313
Db 320 TARGVAVPGHDHYHFTPYSQMSELEERIALIPLVRSNHWVPDSRPEQSPQPTPEPSP 379
QY 314 -----ISGGSVSTNAX-----PNEVYSSLSGSLSNPS 342
Db 380 GQOPAPNLKIDNSLSVLQVKGEGYVFEKGISRYVFAKDLPSFTVKNLESKSKQE 439

QY 343 S-----TTTSKELSSASDGYIFNPK-DIVEETATAVIVRHG--DHFHYIPK-----SNQI 389
Db 440 SVSHTLTAKKNVAPRDQEFYKAYNLLTEAHKALFXNKGNSDFQALDKILERLNDEST 499
QY 390 GQPTLPNNSIATPSP-SLPINPCTSHEKE-----EDGYCFDANRIIA 431
Db 500 NKEKLVDDLLAFAPITHPERLGKPNQSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIIS 559
QY 432 EDESGFVMSHGDNHNYFFPKDLTEQIKAAQKHLSE-----VKTSHNGLDLSLS 480
Db 560 DEGDAYVTPHMGSHWIGKSDLSDEKVAQAAYTREKGLTPSPDADYKANPTG-DSAAA 618
QY 481 HEQDYPGNAKEMKDLDKIEEKIAGIMKOYGVKRESIVVNKKNKNAIIPHGDDHHADP-- 538
Db 619 IYNRVKG-----EKRIPLVRLPYMV--EHTVEKNGNLII-PHKDHYHNKFA 663
QY 539 -IDEH---KPVGIG---HSHSNYELFKPEEGVAKKGNKVTGDELTVNVLNKNSTFN 590
Db 664 WFDDHYRAPNGYTTLEDLFATIKYVVEHDERPHSNDG-----WGNASEHVIGKRDHSEDP 719
QY 591 NQNF 594
Db 720 NKNE 723

RESULT 4
Q97QM8
ID Q97QM8 PRELIMINARY: PRT: 802 AA.
AC Q97QM8;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson R.J.,
Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
McDonald L.A., Feldblyum T.V., Angioli S., Dickinson T., Hickey E.K.,
Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007418; AAK75284.1;
DR TIGR; SP1175;
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1833 CRC64;

Query Match 23.6%; Score 1255; DB 16; Length 802;
Best Local Similarity 41.3%; Pred. No. 2.7e-51;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQREGIOAEQIVIKITDQGYV 60
Db 6 CSYELGLYQARTVKNNRVSYIDGKQATKNTENLTPEVSKREGINAEQIVIKITDQGYV 65
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIYNEVKGGLIKVDGKYVYLKD 120
Db 66 TSHGDHYHYNGKVPYDAIISEELLMKDPNYKLKDEDIYNEVKGGLIKVDGKYVYLKD 125
QY 121 AAHADNVRTKDEINRQKQEHVKDNE----KVNSNVAVARSQGRYTTNDGYVFNPAIDIED 176
|||||

Db 126 AAHADNVRTKEEINRQKQEHSHQREGGTPRNDGAVALLARSQGRYTTDDGYIFNASDIIED 185
QY 177 TGNAYIVPHGHHYHYIPKSLASASLAAAHLAGKNQPSQLSYSTASDN----- 228
Db 186 TGDALIVPHGHHYHYIPKSLASASLAAAHLAGKNQPSQLSYSTASDN----- 245
QY 229 -----NTQSVAGKSTSPANKSENQLSLKELYDPSAQRYSESGLVFPDPAKIIER 280
Db 246 SVSNPGTNTNTSNNTSNQASQNDISLLKOLYKPLPSQRHVESDGLVFPDPAQITTS 305
QY 281 TPNGVAIPHGHYHYIPYKSLASALEKATMVP----- 313
Db 306 TARGVAVPHGHHYHYIPYKSLASALEKATMVP----- 365
QY 314 -----ISGTGSTVSTNAK-----PNEVSSLSGLSSNPS 342
Db 366 GPQAPNLKIDSNLSQVLYKVGEGYVPEEKGISRYVFAKLPSETVKNLESKLSKOE 425
QY 343 S-----LTSKELSSASDGYIFNPK-DIVEETATAYIVRHG--DHEHYIPK-----SNQI 389
Db 426 SVSHLTAKKNWAPROGFEYDKAYNLLTAHKAFLPENKGRNSDPQALDKLLERLNDEST 485
QY 390 GOPTLPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDAARIITA 431
Db 486 NKEKLVDDLLAFLAPITHPERLGRPNQIEYTEDVRIAGLADKYTSDGYIFDEHDIIS 545
QY 432 EDESGFVMSHGDHNYHFKKDLTEQIKAAKHLKE-----VKTSHNGLDLSLS 480
Db 546 DEGDAYVTPHMGHSHWICKDLSKDKVAQAAYTKKGIILPPSPDADYKANPTG-DSAAA 604
QY 481 HQDYPGNNAKEMKLDKIEKIAKIMQYGVKRESIVVKNKNAIYPHGDHHDHP--- 538
Db 605 IYNRVKG-----EKRIPLVRLPMV--EHTVEVKNGNLII-PKHGDYHNIKFA 649
QY 539 -IDER---KPVGIG---HSHSNVLEKFPKPEGVAKKGNKYVTGEELTNVNLLKNSTEN 590
Db 650 WEDDHYKAPNGYLTDLFATIKYIVVHPDHPHSNDG-----WGNSEHVLGKDKHSEDP 705
QY 591 NONF 594
Db 706 NKNF 709
RESULT 5
Q37QM9
ID Q97QM9 PRELIMINARY; PRT: 819 AA.
AC Q97QM9;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Conserved domain protein.
GN Sp1174.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group: Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Heidelberg J., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Desoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA McDonald L.E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDaniel D.A., Felgulyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Wang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-505(2001).
DR EMBL; AE007418; AAK75283.1;
DR TIGR; SP1174; "

KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;
Query Match 23.2%; Score 1235.5; DB 16; Length 819;
Best Local Similarity 41.0%; Pred. No. 2.3e-50;
Matches 297; Conservative 102; Mismatches 183; Indels 143; Gaps 24;
QY 1 CAYALNHRHS--QENKNNRYSYVDGSSQSKSENLPDQVQSKREGIAEQIVIKITDQGY 59
Db 20 CSYELGRYQAGQDKKESNRVAYIDGQAGQAKENLPIDEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGVYDALFSEELLKDPNYOKADIVNEVKGYIILKVGKYYVYLK 119
Db 80 VTSBGDHYHYNGVYDALFSEELLKDPNYOKADIVNEVKGYIILKVGKYYVYLK 139
QY 120 DAAHADNVRTKDEINRQKQEHVD-NEKVNNSVAVARSQGRYTTDDGYIFNASDIIEDTG 178
Db 140 DAAHADNVRTKEELKQKQKQSHHNSRADNAAVAAAGRYTTDDGYIFNASDIIEDTG 199
QY 179 NAYIVPHGHHYHYTPKSDLSASLAAKAHLAGKNQPSQLSYSTASDNNTQ---SVAK 235
Db 200 DAYIVPHGHHYHYTPKSNLSASLAAAEIWNCK--QGRSPSSSYNANPAQRLSENH 257
QY 236 GSTSKPA---NKSENQSLKELYDPSAQRYSESGLVFPDPAKIIISRTNGVAYPHGH 292
Db 258 NLVTPTVYHQNGENISLLRELYAKPLSERHVESDGLIFDPAQITSTRARGVAVPHGNH 317
QY 293 YHFTPYKSLALEEKIARMVPISTGSTVSTNAKN-----EVVSSLSGLSSNPSL 344
Db 318 YHFTPYQMSLEARIARIIPLYRSNHWVDPSPPEPSPQPTPEPSPQAPSNPIDE 377
QY 345 TTSKE-LSSASDGYIFN-----PKDIVEETAT---AYIVRHGDHPHY-----IP 384
Db 378 KLVEARVKGVDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGTKTDLIP 437
QY 385 KSN-----QIGOPTLPNNSLATSPSL----- 406
Db 438 SSDFEFYNKAYDILLARLHQDLLONKRGQVDFEALDNLKDVSSDKVKYKLVEDLLAEIA 497
QY 407 PI-----NPGTSHKHE-----EDGYGFDAARIIAEDESFGVNSHGDRN 445
Db 498 PIHPERLGRPNAGITVTDDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHMTHS 557
QY 446 HYFCKDLTEEQIKAAQKHLKEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLK-----KTE 500
Db 558 HWIKDLSERAEAAQAAYAKE-----KGUTPPSTDHQD--SGNT-EAKGAEAIYNRVKAA 610
QY 501 EKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPIDEHKPVGIGSHSNYELFKPE 560
Db 611 KKVPLDRMPYNLQ---VTVEVKNGSLIIPHVDHYH-----NIKFEW---D 650
QY 561 EGVAKKEGNVYTGEELTNVNLL-----KNSTFNQNFLL---ANGOKRVSEFSP 608
Db 651 EGL--YEAPKGYLTDLATVYVVEHPNRPDSDNGFNASHDVORNGKNGQADINQTEK 708
QY 609 PELEK 613
Db 709 PSEK 713
RESULT 6
Q9ANY2
ID Q9ANY2 PRELIMINARY; PRT: 839 AA.
AC Q9ANY2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHFD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group: Lactobacillales;
OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.;
 RT "Identification and characterization of a novel family of pneumococcal
 RL proteins (the Pht family) that are protective against sepsis.";
 RN Infect. Immun. 69:949-958(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Umayan A.S., Winn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Madan A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey J.C.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*
 RT *pneumoniae*.";
 RL Science 293:498-506(2001).
 DR EMBL: AF318955; AAK06760.1;
 DR EMBL: AE007403; AAK78120.1;
 DR TIGR: SP1003;
 KW Signal; Hypothetical protein; Complete proteome.
 FT SIGNAL
 FT NON_TER 839 839
 SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;
 Query Match 23.2%; Score 1232.5; DB 16; Length 839;
 Best Local Similarity 42.8%; Pred. No. 3.1e-50;
 Matches 295; Conservative 90; Mismatches 164; Indels 140; Gaps 25;
 QY 1 CAYALNQHS-QENKDNRRVYVDSQSQSKSENLTDPQVSKQEGIAEQIVIKITDQGY 59
 DB 20 CSYELGRHQAGQVKESNRVYIDGDAQQAENLTDPDEVSKREGINAEQIVIKITDQGY 79
 QY 60 VTSBGHDHYHYNGKVPYDALFSEELLMDPNYQKLDADIVNEKGGYTIKVDGKYVYVYLK 119
 DB 80 VTSBGHDHYHYNGKVPYDAIIEELLMDPNYQKLDADIVNEKGGYTIKVDGKYVYVYLK 139
 QY 120 DAAHADNVRTKDEINRQKQEHVKD--VAVASQGRYTTNDGYVFNPAIDIED 176
 DB 140 DAAHADNVRTKEIKRQKQEH--SHNNGGSNDQAVVAARAQGRYTTDDGYIFNASDIIED 198
 QY 177 TGNAYIVPHGHDHYHYIPKSDLSASELAALAHLAGKNMQPSQLSYSSTASDNTQ---SV 233
 DB 199 TGDAYIVPHGHDHYHYIPKSNLSASELAALAEAYWNGK--QGSRPSSSSSYNANPAQPRLSE 256
 QY 234 AGSTSKPA---NKSENLOSLLKELYDSPAQRYSESGLVDFDPAKIISRTNGVAIPHG 290
 DB 257 NNLVTPTVYHQNGENISSLLRELKPLSERHVESDGLIFDPQAITSRTARGVAVPHG 316
 QY 291 DHVFTIPYKLSALEKIAARMVPISTGTSTVSTNAKPN-----EVVSSLSGS 336
 DB 317 NHYHFTIPYQMSLEKRIARIIPLYRSNHWVDSRPEQSPQSTPEPSPQAPNPQP 376
 QY 337 LSNPSSLTTSKE-LSSASDGYFN-----PKDIVEETAT---AVIVRHGDHFHY- 382
 DB 377 APSNPIDEKLVKAEVRKVGDFEENGYSRYPKADLSAETAAAGIDSKLAKQESLSHKL 436
 QY 383 -----IPKSN-----QIGQPTLPN-----NSLA-----TPSPSL--- 406
 DB 437 GAKKTDLPSDREFYFNKAYDLLARHQDLDLANKQVDFEALDNLRLERLKVPSDKVLY 496
 QY 407 -----PI-----NPGTSKEHE-----EDGYCFDANRIIAEDSGF 437
 DB 497 DDLAPLAPIRHPERLGRKPNQAITTDDIEIQVAKLAGKYTTEDGYIFDPRDITSDGDAY 556

QY 438 VMSHGDHNYFFKKDLTETEQIKAAQKHLEEVKTSHNGLDLSLSSHEQDYFGNAKEMKDLDK 497
 DB 557 VTPHMTSHWIKKDSLSAEARAAQAYAKE-----KGLTPPSTDHQD-SGNT-EAKGAE 609
 QY 498 -----KIEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDHPIDEHKPVGISHSHS 552
 DB 610 IYNRVKAARKVPLDRMPYNLQ---YTVVEKNGSLIIPHYDVH-----NI 651
 QY 553 NYELFKPEGVAKKEGKNVYTGEEELTNV 581
 DB 652 KEWF--DEGL--YEAPKGYTTLELLATV 676

RESULT 7

Q9ANV3 PRELIMINARY; PRT; 819 AA.
 AC Q9ANV3.
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (Fragment).
 GN PHTB.
 OS *Streptococcus pneumoniae*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.;
 RT "Identification and characterization of a novel family of pneumococcal
 RL proteins (the Pht family) that are protective against sepsis.";
 RL Infect. Immun. 69:949-958(2001).
 DR EMBL: AF318954; AAK06759.1;
 KW Signal.
 FT SIGNAL
 FT NON_TER 819 819
 SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 23.2%; Score 1232.5; DB 2; Length 819;
 Best Local Similarity 41.0%; Pred. No. 3.1e-50;
 Matches 297; Conservative 101; Mismatches 184; Indels 143; Gaps 24;

QY 1 CAYALNQHS-QENKDNRRVYVDSQSQSKSENLTDPQVSKQEGIAEQIVIKITDQGY 59
 DB 20 CSYELGRYQAGQDKESNRVYIDGDAQQAENLTDPDEVSKREGINAEQIVIKITDQGY 79
 QY 60 VTSBGHDHYHYNGKVPYDALFSEELLMDPNYQKLDADIVNEKGGYTIKVDGKYVYVYLK 119
 DB 80 VTSBGHDHYHYNGKVPYDAIIEELLMDPNYQKLDADIVNEKGGYTIKVDGKYVYVYLK 139
 QY 120 DAAHADNVRTKDEINRQKQEHVKD--NEKVSNNVAVARSQGRYTTNDGYVFNPAIDIEDTG 178
 DB 140 DAAHADNVRTKEIKRQKQESHNSRADNAVAARAQGRYTTDDGYIFNASDIIEDTG 199
 QY 179 NAYIVPHGHDHYHYIPKSDLSASELAALAHLAGKNMQPSQLSYSSTASDNTQ---SVAK 235
 DB 200 DAYIVPHGHDHYHYIPKSNLSASELAALAEAYWNGK--QGSRPSSSSSYNANPAQPRLENH 257
 QY 236 GSTSKPA---NKSENLOSLLKELYDSPAQRYSESGLVDFDPAKIISRTNGVAIPHGDH 292
 DB 258 NLVTPTVYHQNGENISSLLRELKPLSERHVESDGLIFDPQAITSRTARGVAVPHGNH 317
 QY 293 YHFTIPYKLSALEKIAARMVPISTGTSTVSTNAKPN-----EVVSSLSGSLSNPSL 344
 DB 318 YHFTIPYQMSLEKRIARIIPLYRSNHWVDSRPEPSPQSTPEPSPQAPNPQ 377
 QY 345 TTSKE-LSSASDGYFN-----PKDIVEETAT---AVIVRHGDHFHY-----IP 384
 DB 378 KLVKAEVRKVGDFEENGYSRYPKADLSAETAAAGIDSKLAKQESLSHKLTKTKDLP 437


```
DR EMBL; AF062533; AAD13797.1; -
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match
  16.9%; Score 899.5; DB 2; Length 822;
Best Local Similarity 27.0%; Pred. No. 1.6e-34;
Matches 262; Conservative 144; Mismatches 297; Indels 267; Gaps 31;

QY 2 AYALNQRHSEKNNRVSYVDGSSQSSQS--ENLTPDVQSQKEGIAQAEQIVIKITDQGY 59
DB 22 SYQLGKHGHLATKNOAIYDSDSKVAKPKTNKTDQISAEEGISAEQIVVITDQGY 81
QY 60 VTSHGDDHYHYNGKVPYDALFSEELLKMPNYQLKADIVNEVKGGYIIKVDGKYVYVLK 119
DB 82 VTSHGDDHYHYNGKVPYDAIIEELLMTDPNHFQKQSDVINEILDGYSVIVKNGVYVYLK 141
QY 120 DAHADNVRTKDEINRQKEHVKD--NEKVNNSVA-----VARSQGRYTTNDGKY 167
DB 142 PGSKRKNIRTKQIAEQVAKGTEKEAKGLAQVAHLSKEEVAANVAKRQGRYTTDGYI 201
QY 168 FNPADIETDNAYIVPHGCHYHYIPKSDLSASELAAKAHLA---GKNMPSQLSYST 224
DB 202 FSPDTIIDDLGDAYLVPHGNHYHYIPKDLSPSELAQAAYWSQKGRGARS--DYRPT 259
QY 225 ASDNNTQSVAKGSTSKPAN-----KSENLOSLKEL 255
DB 260 PAPGRKAPIFDVTPNPGQHQPDNGYHPAPPNPNDASQNKHQDEKFGKTFKELLDQ 319
QY 256 YDPSAQRYSSEGLVDFDPAKIIISRTPNQVAIPHGDHYHFIPYKLSALEEKIARWPI 315
DB 320 HRLDLKYRHEEDGLIFEPTQVKSNAFYVPHGDHIIIPRSLSPLEMLADRYLAG 379
QY 316 GTGSTVSTNAKNEVSSLSGSSSPSSSLTTSKELS-----SASD 355
DB 380 QTDDNDS-----GSDHSKPS-----DKEVTHFTFLGHRKAYKGLDGKPYDTS 423
QY 356 GYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPN--NSLATPSPSLP 413
DB 424 AYVFSKSTHSVDKSGVTAKHGDFHYI--GFGELEQYELDEVANVWVAKQGADELVAALD 482
QY 414 HEKHEEDGYGFD---ANRIIAEDESQFVNSHGDHNFHFKDLTEEQTKAAQKHLEEVK 469
DB 483 QEQGKEKPL-FDTKKVSRKVTGKGVYIMPDKDGYFYARYQLDITQIAFAEQELMLK 541
QY 470 TSHNGLDLSLSSHEQDYPGNNAKEMKDLKIEKTIAGIMKQYGVKRESIVVNEKNAIYP 529
DB 542 KKHRYVDI-----VDTGIEPRLAVDVSSLPNHAGNATYDTGSSFVI-P 583
QY 530 HGDHHRADP---IDERKPVGIGHSHSNYELFKPE--EGVAKKEGKNVYTGELTNVNL 584
DB 584 HIDHIHVYPYSLNRQIATI-----KYVMQHPEVPRPDVWSKPGHE--ESGVSIPNVPTLD 637
QY 585 KNSTFNQNFLLANGQKRVSFSPPELEKLGINMLVLTDPDGKVLKYSKGVFGEGVG 644
DB 638 KRAGMFWQI-----IHSABEVOKALAEQ---RFAAPDGYFD-----672
QY 645 NIANFELDQYLPFGOTFKYTIASKDPEYSYDGTFTVPTSLAYKMASQTIFFPFHAGTY 704
DB 673 -----PROVLAKETP-----VWKDGSFSIPA-----694
QY 705 LRVNPOFAVPGTDALRVDFEFGHAYLNNYKYGEIKLPIKLNQGTTRTAGNKIPVT 764
DB 695 -----DGSSLRITINKSDL- 707
QY 765 FMANAYLDNQSTYIYVPTILEKEN-----QTDKPSILPQFKNKQAQBSKLDKVEEPRKT 819
DB 708 -----SQAEWQAQELLAKNAGADATDTPKPEEQQ--ADKSNEN-----QQPSE 750
QY 820 SEKVEKEKLSGTNSTLTBEVPTVD-----PVQEKVAKFAESYCMKLENVLENMGTI 875
DB 751 ASKEEKE-----SDDFIDSLPDYGLDRATLEDHINQLAQKANIDPKYLIFQPEG-V 800
QY 876 ELYLPSGEVI 885

Query Match
  16.9%; Score 897; DB 2; Length 825;
Best Local Similarity 26.5%; Pred. No. 2.1e-34;
Matches 258; Conservative 152; Mismatches 293; Indels 270; Gaps 31;

QY 2 AYALNQRHSEKNNRVSYVDGSSQSSQS--ENLTPDVQSQKEGIAQAEQIVIKITDQGY 59
DB 22 SYQLGKHGMSATKDNQIAYIDDSKGRAPKTNKTDQISAEEGISAEQIVVITDQGY 81
QY 60 VTSHGDDHYHYNGKVPYDALFSEELLKMPNYQLKADIVNEVKGGYIIKVDGKYVYVLK 119
DB 82 VTSHGDDHYHYNGKVPYDAIIEELLMTDPNHFQKQSDVINEILDGYSVIVKNGVYVYLK 141
QY 120 DAHADNVRTKDEINRQKEHVKD--NEKVNNSVA-----VARSQGRYTTNDGKY 167
DB 142 PGSKRKNIRTKQIAEQVAKGTEKEAKGLAQVAHLSKEEVAANVAKRQGRYTTDGYI 201
QY 168 FNPADIETDNAYIVPHGCHYHYIPKSDLSASELAAKAHLA---GKNMPSQLSYST 224
DB 202 FSPDTIIDDLGDAYLVPHGNHYHYIPKDLSPSELAQAAYWSQKGRGARSYRPTPA 261
QY 225 ASDNNTQSVAKGSTSKPAN-----KSENLOSLKEL 255
DB 262 PAPGRKAPIFDVTPNPGQHQPDNGYHPAPPNPNDASQNKHQDEKFGKTFKELLDQ 321
QY 256 YDPSAQRYSSEGLVDFDPAKIIISRTPNQVAIPHGDHYHFIPYKLSALEEKIA-RMVPI 314
DB 322 HRLDLKYRHEEDGLIFEPTQVKSNAFYVPHGDHIIIPRSLSPLEMLADRYL-- 379
QY 315 GTGSTVSTNAKNEVSSLSGSSSPSSSLTTSKELS-----SAS 354
DB 380 --AGTETDDDS-----GSDHSKPS-----DKEVTHFTFLGHRKAYKGLDGKPYDTS 424
QY 355 DGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPN--NSLATPSPSLP 412
DB 425 DAYVFSKESIHSVDKSGVTAKHGDFHYI--GFGELEQYELDEVANVWVAKQGADELAAAL 483
QY 413 SHEKHEEDGYGFD---ANRIIAEDESQFVNSHGDHNFHFKDLTEEQKAAQKHLEEV 468
DB 484 DOEQGEKPL-FDTKKVSRKVTGKGVYMPDKDGYFYARYQLDITQIAFAEQELMLK 542
QY 469 KTSNGLDLSLSSHEQDYPGNNAKEMKDLKIEKTIAGIMKQYGVKRESIVVNEKNAIYP 528
DB 543 DKKHRYVDI-----VDTGIEPRLAVDVSSLPNHAGNATYDTGSSFVI- 584
QY 529 PHGHHHRADPI-----DEHKVGVG-HSHSNYELFKPE--EGVAKKEGKNVYTGELTNV 581
```

Db 585 PHIDHIVVPSWLTROJATI-----KYVMQHPEVVRPDINSKPGHE-ESGSVIPNVT 636
Qy 582 NLLKSTFNQNFILANQKRVSPFFPELEKKGILNMLVKLITPDGKVLKSVKQVGE 641
Db 637 PLDKRAGPNWOI-----IHSAEVOKALAEG---RFAIPDGVIFD----- 674
Qy 642 GVGNIANFELQPLPGOTFKYITASKDYPEVSDGTFTVTPTSLAYKNASOTIIFYPFHAG 701
Db 675 -----PDVLAKETF-----VWKDGSFIPRA----- 696
Qy 702 DTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKI 761
Db 697 -----DGSRLRTNKS 707
Qy 762 PVTMANAYLDNOSTYIIVEPILEKEN-----QTDKPSILPQFRKNAQENSKLDERVEE 816
Db 708 DL-----SOAEWQAQAEALLAKKNAGDATDIDKP-----KEKQADKSNENQOP 750
Qy 817 PTKSEKVEKELSETGNSNSTLEEVPTVD-----PVOEKVAKAEVSGMKLENVLEND 872
Db 751 SEASKEEKE-----SDDFIDSLPDYGLDRATLEDHINQAQKANIDPKYLIFQPE 801
Qy 873 GTIELVLPSEGEVI 885
Db 802 G-VOFNKNGEIV 813
RESULT 11
Q9AE21 PRELIMINARY; PRT; 289 AA.
ID Q9AE21
AC Q9AE21
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCB_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBSII, an active group
II intron identified in human isolates of group B streptococci.";
RL J. Bacteriol. 183:2560-2569 (2001).
DR EMBL; AJ290952; CAC35985.1; -
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;
Query Match 12.4%; Score 661.5; DB 2; Length 289;
Best Local Similarity 50.6%; Pred. No. 6a-24; Mismatches 66; Indels 19; Gaps 5;
Matches 130; Conservative 42;
Qy 2 AYALNHRSOENKDNRYSDVGSQSQRK--ENLTPDQVSKQEGTQAEQIVKIKITDQGY 59
Db 22 SYQLGKHMHGLATKQIAVDDSKGKVRKPKTKMTMDQISABEGISAEQIVVKITDQGY 81
Qy 60 VTSBGHDHYHNGKVPYDALFSELLMKDNYQLKADIVNEVKGYYIIVKDGKYYVYLK 119
Db 82 VTSBGHDHYHNGKVPYDAIISELLMTDNYRFGKSDVINEIDLDGYIVKNGNYVYVYLK 141
Qy 120 DAHADNVRTKDEINRQKQEHVKD-NEKVNVA-----VARSGRYTTNDGVV 167
Db 142 PGSRRKRIKQQLAEQVAKGTKEAKGLAQVAHLKSKEEVAVNEAKRGRYTTDDGYI 201
Qy 168 FNPADIETDGNAYIVPHGGHYIIPKSDLSASELAHAAKLAH---GKNMQPSQLSYSST 224
Db 202 FSPDTIIDLDGDAYLVPHGNHYIIPKDLSPSELAAQAYWSQKQGRGAPRS--DYRPT 259
Qy 225 ASDNNTOSVAKGSTSKP 241

Db 260 PAPGRKRAPIPDVTPNP 276
RESULT 12
Q99276 PRELIMINARY; PRT; 792 AA.
ID Q99276
AC Q99276
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DE Putative intercalin A precursor.
GN INLA OR SPV1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCB_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kerton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL; AE006574; AAK34188.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_Out.
DR Pfam; PF00560; LRR; 4.
DR SMART; SM00370; LRR; 4.
KW Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32288485ACE0 CRC64;
Query Match 5.5%; Score 292.5; DB 16; Length 792;
Best Local Similarity 21.4%; Pred. No. 6.2e-06;
Matches 185; Conservative 113; Mismatches 285; Indels 281; Gaps 45;
Qy 217 SOLSYSTASDNN-----TQSVAKGSTS---KPAKSENLSQLLKEL---YDPSQAQRY 265
Db 16 SOLTILACQSRNGTYPIKTKQSRKGMTSNKIKIKKSKTKTKTKGVAGVDEPT----- 70
Qy 266 ESDGLVF-DPAKILSRTPNGVAIPHGHYHPIVPSKLSALEEKIARMVPISGTGVSTN 324
Db 71 -DDGFILTKDSKILSKTKDQGLIVDHDGHSHEIFYADLKG--SPEYILIP---KGASL--- 121
Qy 325 AKPNEVVSIGLSLSPSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGHDHFHYIP 384
Db 122 AKP-----AVQRAASQGTSK-VADPHHYEFNPADIVAEADALGYTVRHDHFEVIL 172
Qy 385 KSNIGQQTLPNNSLAT--PSPSLPINPTSH-----EKHEEDGYGDANRIAEDESGF 437
Db 173 KSSLSGQTAQAKQAVATRLPQTSSLVSTATANGIPGLHFTSDGFQFNGQGVGVTKDSI 232
Qy 438 VMSHGDHNYHFFKDLTE-----EQJKAQKHLVEEVTSHNGLDLSLSHEQDYPG 487
Db 233 LVYDGHCHLHPISFADLRGGWAHVADQYDPAKAEKPAETHQT----- 275
Qy 488 NAKEMKDKDKIEEKIAGIMQYGVKRESI---WYNKEKNATIIYPHGDH----- 533
Db 276 --PELSEREKEYQELAYLAELKIDGSTIKRVETQDQKGLGLEYPHHDHVAHMLSDIEI 333
Qy 534 -----HHADDIDHKKPVG-----IGHSHSNYELFKPEEGVAKKE 567
Db 334 GKIDPDPAIHARELEXHK-VGMDTLRALGDFDEVIDIVRTHCAPTFPSNE---KD 388
Qy 568 GNKVTGTELTNWNVL---LKNSTFNQNFILANQKRVSPFFPELE-----KLGII 617
Db 389 PNMM--KEWLATVIKLDGSRKDPQLRKGSLPLNLETLGIGFTPIKDISPVLFQKRLQ 446
Qy 618 NMLVKLITPDGKVLKSVKQVGEVGNIANFELDQPLPGQTFKYTTASKDYPEVSYDG 677


```
Db 447 LLMTKGTVDYRFLD-----NNPQLEG-----IDISQNNLKDISELS 483
Qy 678 TTFVPTSLAYKASQTIFYPPHAGTYLR-VNPQFAVPKGTDALYRVDFEFGNAYLENN 736
Db 484 KYKNTLVA-----AANGIEDIRPLGOLP-NLKFV-----LNN 518
Qy 737 YKVGELKIPKLNQGTTRTAGNKIPVPMANAYLDNOSTYIVEV-PILEKENQT-----790
Db 519 -KISDLS-PLASLHQ-----LQELHIDNQ--ITDLSPVSHRESLIVVDLS 560
Qy 791-----DKPST-----LQFKKKAQENSKLDEKVEEPT 819
Db 561 RNADVLTATQAPKLETLMVNDTKVSHLDLKNPNLSLSINRAQLSL--EGTEASSV 618
Qy 820 SEKVEKEKLSGTNSTSTLEVEPTVDVQEVKAKFAESYGMKLENVFNMDG-----873
Db 619 IVRVAE-----GNIKSLVKD-----KQSLTFLDVIG-----NOLTSLEGVNNETA 662
Qy 874-----TIELYPSGEV-----TKN-----MADF---TGEAPQNGENKPS--ENG 909
Db 663 LDILSVSKNQLTNVLSKPNKVTWIDISHNNISLADKLINQOHIPEAIKKNFAVYEGS 722
Qy 910 KYSTGTVENOPT-ENKPADSLPEA 932
Db 723 MYGNGTAEKAMATKAKESAQEA 746

RESULT 13
Q97QP7
ID Q97QP7 PRELIMINARY; PRT; 2004 AA.
AC Q97QP7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
GN Sp1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.F., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007416; AAK75263.1; -.
DR MEROPS; M26.001; -.
DR TIGR; Sp1154; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000130; Zn_Mipectase.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXIG_anchor; 1.
DR TIGRFAMS; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004 AA; 223908 MW; 556EC6A1028D60A4 CRC64;

Query Match 4.5%; Score 238.5; DB 16; Length 2004;
Best Local Similarity 22.6%; Pred. No. 0.0081;
Matches 206; Conservative 110; Mismatches 284; Indels 313; Gaps 52;
Qy 186 GGHYHYPKSDLSASELAATAKHAHLAKGNQPSQLSY---SSTASDNNTQSVAKGSTSRPA 242
```

```
Db 47 GVHHKYVADSLSSEE-----KKQLYDIPYVNDDETYLV-----84
Qy 243 NKSENLQSLKELXDSPSAQRYSESQGLVFPDAKIIIRTPNGVAIPHGDHYHFIPYSKLS 302
Db 85 -YKLNQNLAEI--PNTGSKNERQALVAG-----ASLAAMGILI-----FAVSKKK 128
Qy 303 ALEKIAMVPISTGS--TVSTNAKPNENVSSLGSSSPSSLTTSKELSSASDGVIFN 360
Db 129 VKNTVLHLVLAGIGVGVSHALENHL-----LYN-----TYELTSGEK--LPL 176
Qy 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKEHEED 420
Db 177 PKEISGVYIYI-KEGKTTSEVSQ-----KSSVATPT-----KQKV 216
Qy 421 CYGDANRIADSDSGFVMSHGDHNYFFKKDLIEBQIAAKQHLBEVKTSHNGLDLSLS 480
Db 217 DYNTPNFV---DHPSTVQAIQEQTPVSSTKP-TEVOVVEKPFSTELINPRKEEKQSDS 272
Qy 481 HEQYFPGNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVNERKNAIYVPHGDHHPID 540
Db 273 QEQ-----LAHKNLETKEEKIS-----PREKTGV-----NTLNPQD 305
Qy 541 E-----HKPVGIGHSHSNYELFPREGVAKKEGNKYVTGEELTNVNLKNSTFNQNF 594
Db 306 EVLSGQLNKP-----ELLYREETME---TKIDFOEI-----QENP 338
Qy 595 TLANGQKRVSPFPPELEKLGINN-LVKLIT-----PDGKYLEKVS 636
Db 339 DLAEGRV-----KQEGKLGKVKVEIVRFSVNKEEVSREIVSTSTAPSRIVEKGT 392
Qy 637 K--VFGE-----GV-----GNIANFELDQYPLPCQTFKYTIASKDYPEVSDGTFTVPT 683
Db 393 KTVIKQEPETGVEHKVQSGAIVEPAI-QPELP-----EAVVSDKGEPEVQ-----TLPE 443
Qy 684 SLAYKMASQTIYFPFHAGDTYLRVNPQ----FAYPKGTDALYRVDFEFGHAYLENNKY 739
Db 444 AV-----TDKGET--EVQPSPTVVSQKGEPEQVAPLPEYKGN-----I 482
Qy 740 GEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSTYIVFPILEKENQDKPSIL 796
Db 483 BOVKPETPVEKTKQGPKEPT--EEVPV-----KPEETPVNPGSTEGTSI- 527
Qy 797 PQFRN---KAQENSKLDEKVEEPEKTSK-----VEKELSETGNST 835
Db 528 -QEAENPVQPAEESTTNSEKV-SPDTSKNTGEVSSNPDSSTTSVSGESNKEPHNDKSNEN 585
Qy 836 SNSTLEEPTVDVQEVKAKFAESYGMKLENVFNMDGTIELYLPSEVYIKKNMADFTGE 895
Db 586 SEKTVEEVP-VNP-----NECTVE-----GTSNQETE 611
Qy 896 APQNGENKPSNGKV---STGTVENQPTENKP--ADSLPEAPNEKPKVPKPS-----TDN 946
Db 612 KPVPFAEETQTNQSKIANENTGETSVSNKPSDSKPPVEESNQPEKNGATKPENSGNTTSEN 671
Qy 947 GMLNPE---GNVGDPMLDPALEAPAVDPVQEKLEKFTASYGLDVSIFNMDGTIELR 1003
Db 672 QCTEPSPSNGNSTEDVSTESNTSNGNEEKQENE-----LDPKKVVEPEKTELELR 724
Qy 1004 LPSGEVTKKNLSD 1016
Db 725 -----NVSD 728

RESULT 14
Q963TI
ID Q963TI PRELIMINARY; PRT; 1078 AA.
AC Q963TI;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).
GN GLURP.
```

OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21313688; PubMed=11420113;
RX Theisen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RT glutamate-rich protein (GLURP) from Plasmodium reichenowi";
RL Mol. Biochem. Parasitol. 115:269-273(2001).
DR EMEL; AF356828; AA40236.1;
FT NON_TER 1 1078
ET NON_TER 1078 1078
SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B223913C CRC64;

Query Match 4.4%; Score 235.5; DB 5; Length 1078;
Best Local Similarity 19.6%; Pred. No. 0.0047;
Matches 209; Conservative 157; Mismatches 380; Indels 321; Gaps 51;

QY 6 NOHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGYVTSBGD 65
DB 200 HEHKKLDNHYEKNVFNHSSNSYNQENLKPSFDEHLNISK-----KLEDOLDLHEHGN 255
QY 66 HYHYNGKVPYDALFSEEL-LMKDNVYL-----KDADIVNEVKGYYLIK 109
DB 256 SEHLKDEKIGNPLVHENSIFNDIEQILNLPQETNVQEQLYNEKQNVKPEKPNSEIFS 315
QY 110 VPGKYYVYLKDAADNADNRTKDEINRQKQ-----DNEKVN---148
DB 316 LD-----LKETTNDILPNQPLENIKQSESEINHVDHALPKENIIDKLDNKEHEHDE 369
QY 149 --SNVAVARSQGYTTNDGYV-----FNPADI-----IEDTGNAYIVPHGG 187
DB 370 SQHNVNVLGN---NINNQLPQEKANIESEFKNIDSEIILPENVEKEIIVDPVSPK 426
QY 188 HY-HYIPKSDLSASELAALAKHAGKNMOPQSLSYSTASDNNTOSVAKGSTSKPANKSE 246
DB 427 HFHNETLEQETSESEHEEA---VSEKNAHETVEHEEAVSQESNPE-----KADNDGNYSQ 478
QY 247 NLOSLLKELYDSPAQSYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYKLSALEE 306
DB 479 NSNNELNE-----NEFVESKSEHEPAE-----NEESSLEE 509
QY 307 KIARVPISGTGSTVETNAKNEVSSLSGSLNSPSSLTSSKLSASDGYTFNPKDIVE 366
DB 510 -----AHQEEIVPEQNQESGESKLVNDNDE-----GGF-----E 538
QY 367 ETATATVVRGHDHFYIPKSN--QIQOPLPNNSLATPSPSLPIPGTSHKEHEED--- 420
DB 579 QNKQESGESKLVNDDEGFEAEHNFSEESNSLHEHEEVESDESDEPPEHEEVSE 638
QY 455 EEOIKAAQKHLEEVKTSNGLDLSLSHEDYPCNAKEMKDLKKTEEKTAGI-----MKQY 510
DB 639 ESNPEASENDESSIEEAHQ--EEISEQNDSELNELVESEKSVSEPAEHVEIVSEKSV 696
QY 511 GVKRESIVNKKENAIITYPHGDHHDADPIDEH---KPVGTGCHSHSYELFKPEEGVAKKE 567
DB 697 SEPAEHVEIVSEKSV---SEPAEHVESVSERASPSSEHVESVSQSNNPESE---KKD 749
QY 568 G---NKYVTGEEITNVNLLKNTFNQNT---LANGQRVSEFSF-----PPELEKLG 617
DB 750 GPVPSKPFEEIEKVDQPKIVDLQIIIEPNFVLDQNPQEPVPSFVKIEKPVSEENKHG- 808
QY 618 NMLVKLITPDGKYLEKVSQKGVFGVGGVGNATANFELDOPILPGQTFKVTIASK-----YEE 672
DB 809 -----SVDPEVEEK-----ENVSEV-----VEKQNPQELFEELPLKDKIEKIEPE 850
QY 673 VSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTALVRFVDFEHCNAY 732

DB 851 LEPEDVHTEQDLLEHK-----TVDPETII-----EVEEIPSELHENEV 887
QY 733 LENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNOSTY-----IVEV-PILEKE 787
DB 888 A--HPEIVEIEVEFPEPQ-----NNEFP-----EINEDDKSAHIQHEIVEVEELPEE 934
QY 788 NQTDK-----PSTLPQFKRKAQENS--KLDEKVEBPKTSEKVEKELKSETGNTSN 837
DB 935 DKNEKVEHEIVEVEEILPENKNEKVEHEIVEVEEILPENKNEKVEHE-----983
QY 838 STLEEYPTVDVQEVAKFAESYGMKLENVL--FNMDGTIELYLPSEGVIKKMMADFTGE 895
DB 984 --IVEVEEILPEEDKNEK-VEHEIVEVEEILPENKNEKVEHEIVEVEEILPEIIEIE-E 1039
QY 896 APOGNGENKPSNGKYSTGTVENQPTENKPADSLPE--APNEKPKVPK 940
DB 1040 VPQGTNNENIETIK-----PEEKNEFSVVEEKAIPQEPVVP 1077

RESULT 15
O96134 PRELIMINARY; PRT; 2485 AA.
ID O96134;
AC O96134;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ser/Thr protein kinase.
GN PFB0150C
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravid L.,
RA Koonin E.V., Shalitin H., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perle M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Science 282:1126-1132(1998).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMEL; AE001376; AAC71820.1;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 2485 AA; 293764 MW; 4297C5CFF030AD7E CRC64;

Query Match 4.3%; Score 228; DB 5; Length 2485;
Best Local Similarity 18.1%; Pred. No. 0.034;
Matches 208; Conservative 149; Mismatches 352; Indels 400; Gaps 47;

QY 3 YALNCHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGYVTS 62
DB 564 YKKRKSNNNNNNNNNIS--SSSSSSSKKNHVI---INNK-----ISS 601
QY 63 HGHDHYHYNGKVPYDALFSEELM-----86
DB 602 YNHYKERKDFKFNELFKKILPKSKDTCVFNERQKDLFEKSENEHICVSNFNTSDD 661
QY 87 -----KPNYOLKADIVNEVKGYYIKTVGKYYVYLKDAADNADNRTKDEINRQK 137
DB 662 ISSHVSNNKKEPFFALKNNSTRHLPKNNIITYTSKSNHVDKREKIVLLKKKEIN--- 718
QY 138 QEHVKNDEKNSNVAVARSQGYTTNDGYVFNPAI-----IEDTGNAYIVPHGGHY 189
DB 719 DKNTSSCLINNIIT-----YTLQNGVNKNLMLGRDTSYIKDEKNMLKCYNGNN 772

Search completed: May 13, 2003, 13:56:33
Job time : 55.2509 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:52 ; Search time 15.7951 Seconds
(without alignments)
1898.176 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315

Sequence: 1 CAYALNHRSENKNNRVS.....TELPLSGEVIKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2338	44.0	447	US-08-961-083-182	Sequence 182, App
2	1247	23.5	796	US-08-961-083-56	Sequence 56, Appl
3	1228.5	23.1	763	US-08-961-083-66	Sequence 66, Appl
4	257	4.8	10182	US-09-134-001C-3159	Sequence 3159, Ap
5	223	4.2	1964	US-08-790-912-3	Sequence 3, Appl1
6	223	4.2	2052	US-08-790-912-2	Sequence 2, Appl1
7	205.5	3.9	1183	US-08-447-031A-2	Sequence 2, Appl1
8	202	3.8	1848	US-08-296-791-6	Sequence 6, Appl1
9	202	3.8	1848	PCT-US93-10661A-6	Sequence 6, Appl1
10	200	3.8	571	US-08-961-083-4	Sequence 4, Appl1
11	198	3.7	851	US-09-071-035-326	Sequence 326, App
12	198	3.7	851	US-09-071-035-330	Sequence 330, App
13	198	3.7	851	US-09-071-035-334	Sequence 334, App
14	188.5	3.5	3696	US-09-134-001C-5080	Sequence 5080, Ap
15	186	3.5	1231	US-08-904-263A-4	Sequence 4, Appl1
16	185.5	3.5	1507	5268270-2	Patent No. 5268270
17	178.5	3.4	2465	US-08-596-291-3	Sequence 3, Appl1
18	178.5	3.4	2465	US-09-100-804-3	Sequence 3, Appl1
19	177	3.3	1702	US-08-296-791-5	Sequence 5, Appl1
20	177	3.3	1702	PCT-US95-10661A-5	Sequence 5, Appl1
21	175.5	3.3	2466	US-09-080-855-12	Sequence 12, Appl
22	175.5	3.3	2466	PCT-US94-03943-2	Sequence 2, Appl1
23	174.5	3.3	2485	US-09-290-640-46	Sequence 46, Appl
24	173.5	3.3	1177	US-09-134-001C-5106	Sequence 5106, Ap
25	173	3.3	1780	US-08-769-309A-5	Sequence 5, Appl1
26	173	3.3	1780	US-08-994-570-5	Sequence 5, Appl1
27	171	3.2	984	US-08-242-932-2	Sequence 2, Appl1

Sequence 2, Appl1
Sequence 2, Appl1
Sequence 24, Appl1
Sequence 30, Appl1
Sequence 5, Appl1
Sequence 6, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 8, Appl1
Sequence 6, Appl1
Sequence 10, Appl1
Sequence 2, Appl1
Sequence 246, App

ALIGNMENTS

RESULT 1
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-182

Query Match 44.0%; Score 2338; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 6,7e-160;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LNHRSENKNNRVSVDGSSQSKSENITPDQVSQKEGIAEQIVIKITDGYVTSHG 64
Db 1 LNHRSENKNNRVSVDGSSQSKSENITPDQVSQKEGIAEQIVIKITDGYVTSHG 60

Query Match 23.5%; Score 1247; DB 4; Length 796;
Best Local Similarity 41.2%; Pred. No. 3.4e-81;
Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 65 DHYHYNGKVPYDALFSEELLMKDPNOLKADIVNEVGKGYIIKVDGKYVYVYLKDAHA 124
DB 61 DHYHYNGKVPYDALFSEELLMKDPNOLKADIVNEVGKGYIIKVDGKYVYVYLKDAHA 120
QY 125 DNVRYKDEINROKQEHVDKNEKXVNSNVAVASQGRYTTNDGVYVNPADIIEDTGNAYIVP 184
DB 121 DNVRYKDEINROKQEHVDKNEKXVNSNVAVASQGRYTTNDGVYVNPADIIEDTGNAYIVP 180
QY 185 HGCHYHYIPKSDLSASELAHAHLAGNMOPSQLSYSTASDNNTQSVAGKSTSKPANK 244
DB 181 HGCHYHYIPKSDLSASELAHAHLAGNMOPSQLSYSTASDNNTQSVAGKSTSKPANK 240
QY 245 SENLOSLKELYDSPAQRYSDGLVDFDPKAIISRTNGVAIPHGDIHYFIPYSKLSAL 304
DB 241 SENLOSLKELYDSPAQRYSDGLVDFDPKAIISRTNGVAIPHGDIHYFIPYSKLSAL 300
QY 305 EKIARMVPISTGTSTVNAPNEVWSLGLSSNPSSLTSTKSELSSASDGYIFNPKDI 364
DB 301 EKIARMVPISTGTSTVNAPNEVWSLGLSSNPSSLTSTKSELSSASDGYIFNPKDI 360
QY 365 VETATAYIVRGDHPHYIPKSNQIQPTLPNNSLATPSPILPNPSTSHKHEEDGYGF 424
DB 361 VETATAYIVRGDHPHYIPKSNQIQPTLPNNSLATPSPILPNPSTSHKHEEDGYGF 420
QY 425 DANRIIAEDESFGVMHSHGDNHYFFKK 451
DB 421 DANRIIAEDESFGVMHSHGDNHYFFKK 447

RESULT 2
US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match 23.5%; Score 1247; DB 4; Length 796;
Best Local Similarity 41.2%; Pred. No. 3.4e-81;
Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 2 AYALNHRSEKNDNRVSYVDSQSQSKSENLTPOVQCKEIQAEQIVIKITDGYVT 61
DB 1 SYELGLYQARTVRENRRVSYIDGQATKTEINTPDVSKREGIAEQIVIKITDGYVT 60
QY 62 SHGDHYHYNGKVPYDALFSEELLMKDPNOLKADIVNEVGKGYIIKVDGKYVYVYLKDA 121
DB 61 SHGDHYHYNGKVPYDALFSEELLMKDPNOLKADIVNEVGKGYIIKVDGKYVYVYLKDA 120
QY 122 AHADNVRTKEINRQKQEHVKDNE---KVNSNVAVASQGRYTTNDGVYVNPADIIEDT 177
DB 121 AHADNVRTKEINRQKQEHVREGGTTPRDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
QY 178 GNAYIVPHGCHYHYIPKSDLSASELAHAHLAGNMOPSQLSYSTASDN--- 228
DB 181 GDAYIVPHGCHYHYIPKSDLSASELAHAHLAGNMOPSQLSYSTASDN--- 240
QY 229 ---NTQSVAGKSTSKPANKSENLOSLKELYDSPAQRYSDGLVDFDPKAIISRT 281
DB 241 VSNPGTTNTNTNNSNTNSQASQSDNDIDSLKQLYKLPLSQRHVESDGLVDFDPKAIISRT 300
QY 282 PNGVAIPHGDIHYFIPYSKLSALEEKIARMV--- 313
DB 301 ARGVAVPHGDIHYFIPYSQMSSELEERIIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
QY 314 ---ISGTGSTVSTNAK--- 343
DB 361 POPAPNLKIDSSNSLSQLVRKYCEGVVFEKGISRVYFAKDLPLSETVKNLESKLSQES 420
QY 344 ---LTTSKELSSASDGYIFNPK--DIVETATAYIVPHG--DHFHYIPK-----SNQIG 390
DB 421 VSHTLTAKKENVAPRQGFYDKAYNLLTEAHKALFXKXGRNSDFQALDKLLERLNDESTN 480
QY 391 QPTLPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDANRIIAE 432
DB 481 KEKLVDDLLAFAPITHPERLGRKPNQSOIEYDEVRITLAQLADKYTTSDGYIFDEHDIIS 540
QY 433 DESGFVMSHGDHNYFFKKDLTEREQIRAAKQKHEE-----VKTSHNGLDLSLSSH 481
DB 541 EGDAYVTPHMGHSHWIGKDSLSLDEKVAQAQYTKKEGILPSPDADYKANPTG-DSAAA 599
QY 482 EODYPGNAKENKDKLTKTEKIIAGIMQYGVKRESIVVNKEKNAILYPHGDHHDADP--- 538
DB 600 YNRVKG-----EKRIPLVRLPYVW--EHTVEVKNGLII-PHKDHYHNKFAW 644
QY 539 IDEH---KPVGIG----HSHSNYELFKPEEGVAKKEGNKYVTGELTNVNNLKNSTFNN 591
DB 645 FDDHTYKAPNGYTTLEDLFIATIKYVVEHPDERPHSNDG----WGNASEHVLGKDKHSDPN 700
QY 592 QNF 594
DB 701 KNF 703

RESULT 3
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:

QY 483 QDYPGNAKEMKDLKKI-----EKKTA-----GIMKQGVKRESIVNKKNAIIPHG 532
Db 5676 KDLVNOAKTRIDVAOKLAAAEINISAMENLDRGQNKEDIRSSGAYINADTKVTAYDQA 5735
QY 533 HHADPIDEHKVIGISHSNYELFKPEEGYAKKGNKVYTGEEITNVVNLKNS----- 587
Db 5736 LQNAENIINATP-----VVELNK-----ATIEQALSRRVQAQDLDGVQOLANAKQAT 5784
QY 588 -TFNQNFTLANGQRKVSFPPELEKLGINMLVKLITPGCKVLEKVG-----KV 638
Db 5785 QVINGLIN-SLNDGQR-----ELNLLINSANTRTKVOEELNKATELNHAMEA 5830
QY 639 FGEGVGNANFELDQPLPGQTFKYTIASKDPEV-SYDGTFTVPTSLAYKMASOTIEYP 697
Db 5831 LNSVONVDQVQSSNY-----NEQPEOHNDN-----AVNEAOATI--- 5869
QY 698 FRAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVGEIKL----- 744
Db 5870 ---NNAQPVLDKLAIERLTQVNTTKDALHGAQKLTQDQAAETGIRGLTSLNBPQNA 5926
QY 745 PIPKLNQGTTRAGNKI---PVTFMANAYL-----DNQSTYIVE----- 780
Db 5927 EYAKVTAATREDEVRNIREATLTOTMLGLRKSINKDNKNSKYINEDHDQOQAYDN 5986
QY 781 -----VPILEKENQTDKPSILPQFKRNKAQENSKL--DEKVEEPTSEK---VEKEKLS 830
Db 5987 AVNAQOVIDEQTALSSDTINOLANAVTOAKSNLHGDTKLQHKDQSAKQTTIAQLQNLNS 6046
QY 831 TGNSTNSLIEEPTVDPVQEVKFAESYGMKLENLVNFMMDGTIELY---LPSGEVI-- 885
Db 6047 AQKHMEDSLIDNESTRTOVQOHLTE-----AQALDGLMGALKESIKDYTNIVNSGNYINA 6101
QY 886 ---KKNMADFTGAPOG---NGENKPSN--GKYSTG--TVENOPTENKPADSLPEAPN--E 935
Db 6102 EPSKKQAYDAVQNAQNLINGNPTINKNGVTTATQTKNKDALDGDHRLERAKNNAN 6161
QY 936 KPYKPNSTNDGMLNPEGNVGSDPMDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFN 995
Db 6162 QTIRNLNLNNAQDAEKNL-----VNSASTLEQVQOQNLQ---TAQO---LQNAME 6207
QY 996 MDGTIELR-----LPSGEVIKKNLSDFI 1018
Db 6208 LRSIAKKDQVKADSKYLNEDPOIKQNYDDAV 6239

RESULT 5
US-08-790-912-3
Sequence 3, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939

FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-4U1
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 4.2%; Score 223; DB 2; Length 1964;
Best Local Similarity 22.8%; Pred. NO. 4.3e-07;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

QY 186 GGHVHYIPKSDLSASELAARAKHLAKGNQPSQLSY---SSTASDNNQTSVAKGSTSKPA 242
Db 47 GVHYKYVADSELSSEE-----KKQLVYDIPYVENDDETYLV----- 84
QY 243 NKSENLOSLLKELYDSPSAQRYSESQGLVDPDAKIISRTPNGVAIPHGDHYHIFYPYKLS 302
Db 85 -YKLNQNLAEEL---PNTGSKNERQALVAG---ASLAALGILI-----FAVSKKK 128
QY 303 ALEEKIARMPVISCIGS--TVSTNAKPNVSVSSLSGSSNPSSLLTSKELSSASDGYIFN 360
Db 129 VKNKTVLHLVUAGMGVLSVVALENHLL-----LNYN-----TDYELTSGEK--LPL 176
QY 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIOGOTPLNNSLATPSPSLPINFGTSHKEHEED 420
Db 177 PKEISGVYIGYI-KEGKTTSDFEVSNQ-----EKSAATPT-----KQOKY 216
QY 421 GYGDANRPIAEDSGFVMSHGDHNYFFKKDLTEQIKAAQKHLBEVKTSHNGLDLSLS 480
Db 217 DYNVTPNEV---DHPSTVOAQEOTPVSSIKP--TEVQVVEKPPFSTELINPKKEKQSSDS 272
QY 481 HEQDYPGNAKEMKDLKKIEEKIAGIMKQGVKRESIVNKKNAIIPHGDDHHADPID 540
Db 273 QEQ-----LAHKNLKTKKEKIS-----PKETGV-----NTLNPDQ 305
QY 541 E-----HKPVGIGHSHSNYELFKPEEGYAKKGNKVYTGEEITNVVNLKNSFTNNQNF 594
Db 306 EYLSQQLNKP-----ELLYREETIE---TKIDFQEEI-----QENP 338
QY 595 TLANGOKRVSESFPPLEKKGILNN-LVKLIT-----PDGKYLEKVS 636
Db 339 DLAEQTVRV-----KQSGKLGKKEIVRIEFSVNKEEVSREIVSTSTAPSPRIVEKGT 392
QY 637 K--VFGE-----GV-----GNIANFELDQPLPGQTFKYTIASKDPEVSYDGTFTVPT 683
Db 393 KTVIKKEQPTGVHEHKQVQSGAIVEPAI-QPELP-----EAVVSORGEPEVOP---TLE 443
QY 684 SLAYKMASOTIYFPFHAGDTYLRVNPQ-----FAVPKGTDALRVFDEFHGNAYLENNYKV 739
Db 444 AV-----TDKGET--EVQPESPDIVSDKGEQVAPLPEYKGN-----I 482
QY 740 GEIK--LPIPKL--NOGTTTACNKTPIVTFMANAYLDNOSTYIVFVILEKENQDKPSIL 796
Db 483 EQVKPPTPVKTEKQPEKT--EYFV-----KPTETPVNPNNEGTEGTSI- 527
QY 797 PQFKN---KAQENSKLDEKVEEPTSEK-----EKEKLSSETGNS 836
Db 528 -QEAENPVQPAEESTINSEKVSPTSSSENTGEVSNPSSDTSITVSGESKNKPEHNSKNNS 586
QY 837 NSTLEEVTPVDPVQEVKFAESYGMKLENLVNFMMDGTIELYLPSEGVIKKNMADFTGEA 896
Db 587 EKTVEEVP-VND-----NEGTV-----GTSNQETEK 612


```

QY 897 POGNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPKVPKPS- ----TDNG 94
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 613 PVOPARETQTNISKIANENTGEVSNAPSDSKPPVEESNQPEKNGIATKPSNGTTSNG 672
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |
QY 948 MLNPE 952
| |
Db 673 QTEPE 677
| |

RESULT 6
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-4U1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-2

Query Match 4.2%; Score 223; DB 2; Length 2052;
Best Local Similarity 22.6%; Pred No. 4,6e-07;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps

QY 186 GGHYHIPKSDLSASELAAKAHLAGKNMQPSQLSY---SSTASDNTQSVAKGTSKPA 242
| | | | | : | | | | | : | | | | | : | | | | | : | |
Db 120 GVHYKYVADSELSSSE-----KKQLVYDIPTYVVENDETYLV----- 157
| | | | | : | | | | | : | | | | | : | | | | | : | |
QY 243 NKSENQLSKLLKELYDSPSAQRYSESGLYFDPAKIISRTPNGVAIPHGDHYHFYPSKLS 302
| | | | | : | | | | | : | | | | | : | | | | | : | |
Db 158 -YKLNQCNQLAEI---PNTGSKNERQALVAG-----ASLAALGILI-----FAVSKKK 201
| | | | | : | | | | | : | | | | | : | | | | | : | |
QY 303 ALEEKIARMPVPSGTGS--TVSTNAKPNEVWSLGSLSNPSSLTSTKLSASDGIYFN 360
| | | | | : | | | | | : | | | | | : | | | | | : | |
Db 202 VKNKTVLHLVAVGMGNGVLVSVHALENELL-----LVNV----TDYELTSGEK--LPL 249
| | | | | : | | | | | : | | | | | : | | | | | : | |
QY 361 PRDIVEETATAIVRGDHFHFIPKSNQIQETPLNNSLATPSPSLPNPGTSGHEKHEED 420
| | | | | : | | | | | : | | | | | : | | | | | : | |

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/447,031A
 FILING DATE: 22-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/861,804
 FILING DATE: 21-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/SE91/00707
 FILING DATE: 22-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9003374-7
 FILING DATE: 22-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: McGowan, Malcolm K.
 REGISTRATION NUMBER: 39,300
 REFERENCE/DOCKET NUMBER: 012889-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1183 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-447-031A-2

Query Match 3.9%; Score 205.5; DB 2; Length 1183;
 Best Local Similarity 19.7%; Pred. No. 3.6e-06;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;

QY 16 NNRVYVYDGSOSSQSENLT-PQVQSKEGIAQIVIKITDQGYVTSNGHYHYNGK- 73
 DB 195 NNEKSYV-----SKDITIKQIQGGQQLDLSTLNINVT-----GTHSNYVSGQS 238
 QY 74 --VPYDALFSELLMKDPNQLKADT-----VNEVKGYYIILKVDG---KYVYVYLKRA 122
 DB 239 AITDFEAFPSKIVTNTNTDVTIPQGYSGYNSFSINRYKTIINEOQKEFVNNSQAW 298
 QY 123 HADNVRTKDEINRQKHVDNKEVNSV-----AVARSQGRYTTN 163
 DB 259 YQEH--GKEEVNGKSFNHTVHNANAGIEGTVKGLKVLKQDKTKAPIANVFKPLSK 356
 QY 164 DGYVF---NPADLIETGNAYIVPHGGHYHYIPKSDLSASELAARAKAHLAGKNWQPSOL 219
 DB 357 DGVVKNQKKEIITDANGI-----ANKALPSGDYILKEIEAPRYTFDRDRE--- 406
 QY 220 SYSSTASNNQSVAKGSTSKPANKSENLOSLLKELYDSPSAQRYSESGLVDFPAKILIS 279
 DB 407 -YPETMKDIDNQGYYF--TTIENAKAIEKTKDV-----SAQK-----VWEGTKVKV 448
 QY 280 RTPNGVAIPGHDHYHFTPYSK--LSALEEKIARMVPISGTIVSTNAKPNEVSSLSGL 337
 DB 449 PTTFKLYKQDDNGNTTPVDAEIKKLED-----GTVKVTWSNLPEND----- 491
 QY 338 SSNPSSLTTSKELSSASDGYIFNPKDIVEATAYIVRHGDHFHYIPKS-NOIGQPTLP- 395
 DB 492 -----KNGKAIKYLVKVEVNAQGEDTTPPE 514
 QY 396 -----NNSLATPPSLPI-NPGTSHEKHEEDGYGFDPANRIIAEDSGFVMSGHDHNYFF 449
 DB 515 GYTKENGLVVTNTEKFIETISISGEKVDKQDNQDKGR--PEKVSVNLLANG----- 565
 QY 450 KDLTEECIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKOLDKKIE-EKTAGIMK 508
 DB 566 -----EKVKT-----LDVTSETNWKY-----EFKDLPKYDEGKKTETVT 600
 QY 509 QYGVK-----RESIVYVK-----EKNAIYPHGDHHDADPIDEHKFVGIGHSHSNYELFK 558
 DB 601 EDHVKDYTTDINGTTINKTTPGTSATVTKNWDNNNO--DGKRTEI-----KVELY- 652

QY 559 PEEGVA-----KKEGNKV-YTGEELTNVYVLLKLNSTFNNOQNT 595
 DB 653 -QDKCATGKTAILLNESNNWTHWTGLDEKAGQGVKYTVVEELTKV-----KGYTHVDNDND 707
 QY 596 LANGQKRVSPPELEKKLGINMLVKLITPDGKVLKESKGVFGEG-----VGNIAN 648
 DB 708 M--GNLIVTNKTPETTSISGEKVDKQDNQDKRREKVSVNLLADGKVKVLDVTSETN 765
 QY 649 FELDQPYLP-----GQTFKYTIAS-----KDYPEVSYDGTFTVPTSLAYKMASQTFIYFPHAG 701
 DB 766 WKYEFKDLPKYDEGKKTETVTDHVKDY--TTDINGT-----TITNKYTPG 810
 QY 702 DTYLELV-----NPQFAVPKGTDALYRVFDE-----FHGNAYLENNNYKV 739
 DB 811 ETSATVTKNWDNNQDKRPTKVELYQDGKATGKTAILLNESNNWTHWTGLDEKAKG 870
 QY 740 GEIKLPIPLNQ---GTRTAGNKIPVTFMANAYLQNSTYIVVEPILEKENOTDK---- 792
 DB 871 QQVXYTVVEELTKVKGTYTHVDNNDMGNLIVTNKYTPETTSISGEKVDKQDNQDKRPEK 930
 QY 793 --PSILPQFKENKA-----OENSKLDE--KVEEPTSEKVEKEKLSGTGNT 835
 DB 931 VSVNLLANGKVKTLDTSETNWKYEFKDLPKYDEGKKIEYVTDHVKDYTTDINGTTI 990
 QY 836 SNSTL--EEVPTVD-----PVOEKVAKFAESYGMKLENVLFNMDGTIELYLP 880
 DB 991 TNKYTPGTSATVTKNWDNNQDKRPTKVELYQDGKATGKTAILLNESNNWTHWTG 1050
 QY 881 SGEYIKKNMADFT--GEAPOGNGENKPSGKSVSTGVNENOPTENKPADSL-PEAPNEK-- 936
 DB 1051 LDEKAGQGVKYTVDELTKVNGYTHVDNNDMGNLIVTNKYTPKPKPNRPIYEPKPKDTP 1110
 QY 937 FVRPENS 943
 DB 1111 PTKPDHS 1117
 RESULT 8
 US-08-296-791-6
 ; Sequence 6, Application US/08296791
 ; Patent No. 6245337
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Gene III, Joseph W.
 ; APPLICANT: Falkow, Stanley
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration
 ; TITLE OF INVENTION: Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791
 ; FILING DATE: 25-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:

Db 192 YNNDKY---PAFVRLSGSGTQFIYKKGSRVQLILTEKQGNLLRNMDVGGDNLVLGNA 248
Qy 181 YIVPIGG-----HYHPIKSDLSASELA----- 203
Db 249 YTYGIAGTPTKYVNHENGLIFGNSKEHSDPKGILSDPNTNVAVLGDSGSPFLVVDRE 308
Qy 204 -----AAKHLACKNMQPSQ-----LSYSS-TASDNNNTQSV--AKGSTS 239
Db 309 KGKWLFLSGYDFWAGYNKKSQWENNIYKHEFAEKIYQYQYSGSLIGSNQTYQWQATGSTS 368
Qy 240 KPAENSENLOSLKELYDPSAQR-----SESDGLVDFDPAKILISPTN 283
Db 369 TITGGEPLSDVLDGDKPNHGKSITLKGSTLTLLNNHIDOGAGGLFFEGDYEVKGTSD 428
Qy 284 -----GVAIPHGD-----HYHPIYSKLS-----ALEE 306
Db 429 STTWKAGVSVADGKTVIKVHNPKYDRLAGIKGTLVVEGKNGEGLKVGDGTVLKQ 488
Qy 307 K-----IARVPISGTGVTSTNAK---PNEVWSL--GSLSSNPSSL----- 344
Db 489 KADANKVQAFQGVGIVSGRSTLVLDNDKQVDPNSIYFCFRGRLDLNGSLTFDHIANI 548
Qy 345 -----FTSELSSADGYIFNPKDIVEETATAYIVRHGDHHP----- 381
Db 549 DQGARVNHNTNITNITGESLTNP-----NTITSYNIEAQDDHPLAIRSIPYQOL 603
Qy 382 YTPKSNQIQOPLPNNSLATPSPINPGTSHEKHEEDGYGDA-----NRIAEDES 435
Db 604 YFNQDNR-SYYLLKKG--ASTRSELSPQSGESNENWLYMGRTSDAAKRNVMNHINERNM 660
Qy 436 GPMVSHGDHNYFFFKDLTEEDIKAAQKHLEVKTSNGLDLSLSEHODY--PGNAKEMK 493
Db 661 GF-----NGYF-----GEEETKATQN--GKLNVTFG---KSDQNRFLTGTGTNLNG 702
Qy 494 DLDDKIEEKIAGIMKQYGVKRESIVVNEKNAIYIPHGDDHHADPIDHKKPVGIG----- 548
Db 703 DL-----NVEKGTFLSGRPTPHARDI-----AGISSTKDD 733
Qy 549 -HSHSNYELFKPEEGVAKK-----EGN-KVYTGEELTNVNNLLKNSFTNNQNTLAN 598
Db 734 PIETENNEVVEDDINRNFRKATTNVTCNASLYSGR---NVANITSNITASN-----N 784
Qy 599 GOKRVSFSPPELEKKGILNMLVK-----LITPDGKYLEKVGKVGEGV-GNI----- 646
Db 785 AQVHIGY-----KTGDTVCVRSDYTGTYTCHNSLSEKALNSFNPTNLGRNVNLFTEN 836
Qy 647 ANFELDQYLPQGTQYTIASKDYDEVSVDGTFVTPTSLAYKMASQTIFYPFHAGDTYLR 706
Db 837 ASFTLGKANLFG-----TIQSIGTSQVN-----LKNSHWHLTGNSNVNOLNLTNGHIH 885
Qy 707 VNPQFAPVPGTDALVRVDFEFGNA-----YLENNYKVGELKLPT----- 746
Db 886 LNAQDANKVTYNTLTVNSLSNGSFFYVWDFTNKSNKVVVNSATGNTFLQVADKGTG 945
Qy 747 -PKLNGGT---TRTAGNKIPVTFFMANAYLDN-----OSTYIYEVPILEKENOT- 790
Db 946 EPNHNLTLFDASNATRNRLNLEVT-LANGSVDRGAWKYKLRNVNGRYDLYNPEVEKRNQTV 1004
Qy 791 -----DKPSILPQFKRKAQNSKLDKVEEP----- 817
Db 1005 DTNTITTPNDIQADAPS-----AQSNNEEIAIARVTPPPAPAPATESAIAEQPETR 1055
Qy 818 -----KTSKEVEKELSETGNSSTNLTLEEYPTVDVPOGVKAKFAESYGMKLE 865
Db 1056 PAETAQAMEETNTANSTETAPKSDTATQENPNSVESVPS-----ETTEKVAENPPQENE 1110
Qy 866 NVLFNMDGTIELYPSGEVIKKNMADF-----TGEAPQNGENKPSNGKYSTGTGVENOP 920
Db 1111 TVAKNEQATEFTFQNGEVAKEDQPTVEANTQTNATSEGTETEQT-----TAEKSEP 1165
Qy 921 T-----ENKPADSLPEAPNERPV-----KPENSTDNGLNPEGNVSDP-MLDPALEEA 968
Db 1166 TESVTVSENOPEKTVSQSTEDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAPAEPEV 1225

Qy 969 PAVDPVQOE 976
Db 1226 PTDTNAEE 1233

RESULT 10
US-08-961-083-4
Sequence 4, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 571 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-4

Query Match 3.8%; Score 200; DB 4; Length 571;
Best Local Similarity 22.8%; Pred. No. 3e-06;
Matches 166; Conservative 87; Mismatches 222; Indels 258; Gaps 42;
Qy 361 PKDIVEETATAYIVRHGDHFIYFKSNQIQOPLPNNSLATPSPSLPINPCTSHKHEED 420
Db 17 PRELSGTYTYGTYI-KESGTTSEVSNQ-----KSSVATPT-----KQKV 56
Qy 421 GYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEEDIGKAAQKHLEEVKTSNGLDLSLS 480
Db 57 DYNVTPNFV---DHPSTVQAIQEQTPVSSTKP-TEVQVVEKPFSTELINPRKEEKSSDS 112
Qy 481 HEQDYPGNAKEMKDLKKIEKTAGIMKQYGVKRESIVVNEKNAIYIPHGDDHHADPID 540
Db 113 QEQ-----LAHKNLETKKEEKIS-----PREKTGV-----NTLNPD 145
Qy 541 E-----HKPVGIGHSHSNVELFKPEEGVAKKCKNKKVYTGTEELTNVNNLLKNSFTNNQNF 594
Db 146 EVLSQNLKP-----ELLYREETME---TKIDFOEEI-----QENP 178
Qy 595 TLANGQRKRVSEFPPELEKKGILNMLVKLIT-----PDGKVLKVS 636
Db 179 DLABGTVRV-----KQEGKLGKKEIVIRFVSNNKEEVSREIVSTSTTAPSPRIVEKGT 232

```
QY 637 K--VFGE-----GV-----GNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPT 693
Db 233 KTVIKQPTGVHEKDVQSGAIVEAI-QPELP-----EAVSDKGEPEVQP-----TLPE 283
QY 684 SLAYKMASQTIFFPHAGDTYLRVNPQ-----FAVPKGTDALVRVDFEFGNAYLENNYKV 739
Db 284 AVV-----TDKGET--EVQPESDPTVVSVDKGEPEQVAPLPEYKGN-----I 322
QY 740 GEIK--LPIPKL-NOGTRTAGNKIPVTFMANAYLONQSYIIVEVILKKNQTDKPSIL 796
Db 323 EQVKPEPVEKTEQGEPEKT--EEVPV-----KPTETVPNNEGTEGTSI- 367
QY 797 PQFRN---KAQENSLDEKVEPEPKTSEK-----VEKEKLSETGNET 835
Db 368 -QEAENPVQPAESTTINSEKV-SPDTSSKNTGEVSNPDSSTTSVCSNKPEHDSKNEN 425
QY 836 SNTLEVPVTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPGEVIKKNMADFTGE 895
Db 426 SEKTVEVP-VNP-----NEGTV-----GTSNQETE 451
QY 896 APOGGENKPSGKV---STGTVENOPTENKP--ADSLPEAPNEKPVKPKENS-----TDN 946
Db 452 KPQPAEETQTSKGTANENTGVSKNPKSDSKPPVESNQPKNGTATPENSNGTTSN 511
QY 947 GMLNPE---GNVGSDDPMLDPALEAPVDPVQKLEKFTASYGLGLSDVIFNMDGTIELR 1003
Db 512 GQTEPEPSNGNSTEDVSTESNTSNGNEEIKOENE-----LDPDKKVEPEKTELELR 564
QY 1004 LPSGEVIKKNLSD 1016
Db 565 -----NVSD 568
```

RESULT 11

```
US-09-071-035-326
: Sequence 326, Application US/09071035
: Patent No 6448043
: GENERAL INFORMATION:
: APPLICANT: Gil H. Choi
: TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
: NUMBER OF SEQUENCES: 496
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/071,035
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: A. Anders Brookes
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB369P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 326:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```

```
: MOLECULE TYPE: protein
US-09-071-035-326
```

```
Query Match 3.7%; Score 198; DB 4; Length 851;
Best Local Similarity 19.2%; Pred. No. 7.6e-06;
Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;
```

```
QY 14 KNNRVSYVDGS-----OSSQKSENLTDPQVSKQEGIOAEQIVIKITDQGYVTSHGDIH 68
Db 74 EDNQLRVQOEAVPSVLKSVSSGDLYPD-----GSLIQHG----- 109
QY 69 YNKGVPVYDALFSELL-----MKDPNYQLKDAIVNEVGGYI-IAVD 111
Db 110 YF-----PINGVGNELKGFGRIOILOGSDWEMDNPFI-----SNLFNVVDGYIQLKVN 161
QY 112 GRYVYLKDAAHADAVRTKDEINRQKQBEVDKNEKVNNAVARSQGRYTTNDGVVENPA 171
Db 162 GK---MPSMVSGRISRAPETNPFTTEFESGKETIANLTLIAKFAPELNNDIYT---- 213
QY 172 DLIEDTGNAVIVPHGCHYHYIPKS-----DLSASELAARAAKHAHLAKGNMPSQLSYST 224
Db 214 -----SIOTWLOQSGSYHFFKPKDFEALIDLKNVNSASPAQATPMQSLNVYGSMDRV 268
QY 225 ASDNNTQSAVGSTSKPA-----NKSENLSL-----LKELYDSPAQRYSESGLVDFP 274
Db 269 LOKNNEYAVGISMYSORVGNVEFFGNTENKKGWHTADGMLYLYNQDFAQ-FDEGYWATIDP 327
QY 275 AKLIISRTNGVAIPHGDHYHFTPYSKLSALEEKIARMVYPISTGVSTVSTNAKPNVSWSL 334
Db 328 YRLPGTIVDTRELANGAY-----TGKRSPQSVWVGS-----NNGQVASI 366
QY 335 QSL--SSNPSSLTTSKELSSASDGVIPNPKDIVERETATAYIVRHGDHPHYIPKSNQIQOP 392
Db 367 GMFLDKSNEGNNLVAKSWFLDQOILNGSGITGTTDASITILDNRMHIPPQEVKLNQOG 426
QY 393 TLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDA NR-----IAEDESQFVMSHGDHNYF 448
Db 427 SDKDNSWISLSAAXPLN-----NTGYVFPNSMNTLDVQIERSG---RYGDINEYF 474
QY 449 FKKDITEQIRAAQKHELVKTSNGLDLSLSDHEDYPSGNAKEMKDLKKEEIAGTMK 508
Db 475 VN-----DKTYTNTFAKISK 489
QY 509 QYGVKRESIVVNKEXNAIYPHGDHHDADPIDEHKPVGIGHSHSNYELF-----KPEEVA 564
Db 490 NYGKTVE-----NGTYEYLTVVGXTNEDIA 514
QY 565 KKEGKVVYTGELINVVNLKNSFNNOFTLANGCKRVSFSPPELEKKLGINMLVKLI 624
Db 515 ALSKNKGYT--VLENTANLQAEAGNYVMNTWNNDQEIAGLAYD----- 558
QY 625 TPDGKVLKSVSKVGEQVGN-----IANFELDQYLPQGTFKYTIASKDYPEVSYDGT 678
Db 559 -PMSVISEKIDNGVYRLTANPLQNNASVSTFDPKGL-----EVVAAD-PEISVDQN 609
QY 679 FTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYK 738
Db 610 IITLNSAGLNGSSRSII-----VKTTPEVT-----KEALEKLIQF--OKEHQEKDYT 654
QY 739 VGEIKLPIPKLNOGTT---RTAGNKIPVTFMANAYLONQSYIIVEVILKKNQTDKPSI 795
Db 655 ASSWKVYSEALKQAQTVADQITATQAEVD-QAETELSAVKQLVKVPKE-----VDKTNL 709
QY 796 LPQFKRNKAQENSKLDEKVEPEPKTSEKVEKELSETGNSTSNSTLEEVPVDPQVQKVAK 855
Db 710 LKIITK-----ENKHEQEK-DYTASSWKVYSEALKQ-ACTVADQITATQAEVDQAEAKLRS 762
QY 856 FAEYGMKLENVLFNMDGTIELYLPSEGVTKKNMADFTGEAPQNGENKPKSEN---GKVS 912
Db 763 AVKRLTLK-----NSGENKKE-----QKNGNGNHLNSTGTVDQ 796
QY 913 TGTVENQPTPE-----NKPADSLPEAPNEKPV 938
```

Db 797 TGTKQVPSOGGFRKASQFLPSTGEKXSI 826

RESULT 12

US-09-071-035-330

; Sequence 330, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071.035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: A. Anders Brookes

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB369P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 330:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 851 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-071-035-330

Query Match

Best Local Similarity 19.2%; Score 198; DB 4; Length 851;

Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;

QY 14 KDNRRVSVVDGS-----QSSCKSENLPDQVSQKEGICQAEQIVIKITDQGYVTSFGDHYH 68

Db 74 EDNQRLVQVQEAQSVLKSQSSGGLFDP-----GSLIQHG----- 109

QY 69 YNKGVPYDALFSEBELL-----NKDPNYQLKDAIVNEVKGYYI-IKVD 111

Db 110 YP-----PYNGSYGNELGKFGRIOTILQGSWMENDPNI-----SNLFNVYDKGYLQIMVN 161

QY 112 KYYVYVLKDAADNVRTKDEINQKOEHVKNKNSVAVARSQGRYTTNDGYVFNPA 171

Db 162 GK-----MFSMVSGRSISRAPNTFTTEFESGKETIANLTIKAFAPENLNDIYT----- 213

QY 172 DIIETGNAYIVPGHHVHYTPKS-----DLSASELAAKAHLACKNQPSQLSYST 224

Db 214 -----SIQTLWQQSGSYHYHFKPKRDFEALIDKNVYNNSASPAQATPMQSLNVIYSGMDRV 268

QY 225 ASDNNQTSQVAKGTSKPA-----NKSENLOS-----LKELYDSPAQRYSQSDGLVDFP 274

Db 269 LQKNNEYAVGISMYSQRYGVNTEFCNTENKKGWHTADGMLYLYNQDEAO-PDEGYWATIDP 327

QY 275 AKIISRTNGVAIPGHDHYHPIPYSKLSALEEKTARKVPISTGTGTVSNVNAKNEVVSSL 334

Db 328 YRLPGTVDYTRLANGAY-----TGRKSPQSWYGG-----NNGQVASI 366

QY 335 GSL--SSNPSSLTTSKELSSASDCGYIENPKDIVDEETATAYIVRHGDHFHYIPKSNQIGQP 392

Db 367 GMFLDKSNEGNNLVAKKSWFLDQGIINLGGITGTTDASITILDNRMTHPQEVKLNQ 426

QY 393 TLPNNSLATPSPSLPIPGTSGHEKHEEDGFGFANR-----IIAEDSGFVMSHGDNHNF 448

Db 427 SKDNWSISLSAAAXPLN-----NIGYVFPNSMNTLVDQIEERSG---RYGDINEFY 474

QY 449 PKDLTEBQIAQAQKHLEEVKTSNGLDLSHSEQDYPGNAKEMKDLKKIEBKIAKIMK 508

Db 475 VN-----DKTYTTFAKISK 489

QY 509 QYGVKRESIVVKNKNAIIPHGDDHHPIDEBKPVGIGHSHSNYELF-----KPEGVA 564

Db 490 NYGRTVE-----NGTYEYLTVVGKTNEEIA 514

QY 565 KKEGNKVTGEBELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKKGINMLVKLI 624

Db 515 ALSKNKGYT--VLENTANLQAIKAGNVMKNTWNNDDQEIAGLYAYD----- 558

QY 625 TPDGKVLKSVKVGEGVGN-----IANFELDQVLPQGTQKTYIAISKDYPSYDGT 578

Db 559 -PMSVISEKIDNGVYRLTLANPLONNASVSEFQKGL-----EVAAD-PEISVDON 609

QY 679 FTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPRKGTDALVRVDFEFHGNAYLENNYK 738

Db 610 IITLSAGLNGSSRSII-----VKTTPVET---KEALEKLIQE--QKEHQEKDYT 654

QY 739 VGEIKLPIPKLNQGT---RTAGNKIPVTWMANAYLDNQSTYIVVEPILEKENQTDKPSI 795

Db 655 ASSWKVYSEALKQAQTVAQDTATQAEVD-QAETELRSVAKQVLKVPYTKQ---VDKTL 709

QY 736 LPQKRKAQNSKLDKVEPKTSEKVEKELSETNSTSTLEEVTVDPVQEKVAK 855

Db 710 LKIHK-----ENEXHQEK-DYTASSWKVYSEALKQ-AQTVAQDTATQAEVDQAQAKLRS 762

QY 856 FAESYGMKLENVLFNMDGTTIELYLPSCGEVYIKKNMADFTGEAPOGNGENKPSN---GKVS 912

Db 763 AVKRLTLK-----NSGENKKE-----QKNGNGNHLNTSTGVQ 796

QY 913 TGTVENQPTQTE---NKPADSLPEAPNEKPV 938

Db 797 TGTKQVPSOGGFRKASQFLPSTGEKXSI 826

RESULT 13

US-09-071-035-334

; Sequence 334, Application US/09071035.

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071.035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 334:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-334

Query Match 3.7%; Score 198; DB 4; Length 851;

Best Local Similarity 19.2%; Pred. No. 7.6e-06;

Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;

QY 14 KDNRRVYVDGS-----QSSQSENLTDPQVSKQEGIQAEQIVIKITDQGYVTSHGDIYH 68
Db 74 EDNQLVQVEAVPSVLKSVSGDGLYD-----GSLIQHG----- 109
QY 69 YXNGKVPYDALSEELL-----MKDPYQLKADADIVNEVKGYYI-IKVD 111
Db 110 YF-----PYNGSYGNELLKGFRIQITLOGSDWEMNDPI-----SNLFNVKGYLQLMVN 161
QY 112 GRYVYLKDAHADNVRTKDEINROKQEHVKONKVNNAVARSQGRYTTNDGVVFENPA 171
Db 162 GK-----MPSWVSGRSIRAPENPTTFESGKETIANLTLIAKAPENLRNDIYT----- 213
QY 172 DIETGTNAYIVPHGGHYHYPKS-----DLSASELAHAKHLAGKNMOPSQLSYST 224
Db 214 -----SIQTLOQSSYVHFHKKPRDFEALIDLKNVNASASPAQATPMQSLNVGVMDRV 268
QY 225 ASDNNTQSVAKGSTSKPA-----NKSLENQL-----LKEYLDSPSAORYSESGLVDP 274
Db 269 LOKNEYAVGISMSQRYGVNFEFGTENGKGGHTADGMLYLYNQDPAQ-FDEGVWATIDP 327
QY 275 AKIISRTNGVAIPGDHYHFTIPYSKLSALEBKIAKRVPIISGTGSTVSTNAKPNVSVSL 334
Db 328 YRLPGITVDITRELANGAY-----TGKRSPOSQWVGGS-----NNGQVASI 366
QY 335 GSL--SSPSSLTTSKELSSASDGVIEPKDITVEETATAYIVRHGDHPHYTPKSNQIQP 392
Db 367 GMFLDKSEGNMLVAKSWFLDGOIINLGSIGITTTDASITILDNRMIIHPQEVKLNQ 426
QY 393 TLPPNSLATPSPSLPINTGTSHEKHEEDGYGFANDR-----IIAEDSGFVMSHGDHNYF 448
Db 427 SOKNSWISLSAAXPLN-----NIGYVFPNSMNTLDVQIERSG---RYGDINEYF 474
QY 449 FKKDLTEQIRAAKHLEEVKTSNGLDLSLSEHQDYPGNKAKEMKDKKTEEKIAGIMK 508
Db 475 VN-----DKTYTNTFTAKISK 489
QY 509 QYGVKRESIVNKEKNAYIYPHGDHHDADIDEHKPVGIGHSHSNYELF-----KPEQVA 564
Db 490 NYGKTVE-----NGTYEVLTVGVTNEEIA 514
QY 565 KEGKVVITGELTNVNLKNSTFNNOFTLANGQKRVSFSPPELEKGLINMLVKLI 624
Db 515 ALSKXKGT--VLENTANLQAEAGNYMMNTWNNDQEIAGLYAYD----- 558
QY 625 TPDGVKLVKSVKGFGEVGN-----IANFELDQYLPGOTETTYTASKDYPEVSYDGT 678
Db 559 -PMSVISEKIDNGVYRLFLANPLQNNASVSEFPKGL-----EVVAAD-PEISVDQN 609
QY 679 FTVPISLAYKMASQITVYFFHAGDTYLRVNPQFAVPKGTDLVRVDFDHGNAYLENNK 738
Db 610 IITLASAGLNGSSRSII-----VKTTPVT-----KEALEKLIQE--QKEHQENDYT 654
QY 739 VGEIKLPKLNQGIT---RTAGNKIPYTFMANAYLDNQSYTYVEVPILEKENTDTPSI 795

Db 655 ASSWKVYSEALKQAQTADQTATQAEVD-QAETELRSVAVQLYKVPKTE-----VDKTNL 709
QY 796 LPQFKRNKAQENSKLDEKVEEPTKSEKVEKELSETGTSTNSTNLEEVPTVDPVQEKVAK 855
Db 710 LKIIK-----ENERHOEK-DVTASSWKVYSEALKQ-AQTADQTATQAEVDQAEAKLRS 762
QY 856 FAESYGMKLENVFLNMDGTIELYLPSEGEVIRKKNMADTGTGEPQNGENKPKSEN---GKVS 912
Db 763 AVKRLTLK-----NSGENKKE-----QKNGGNHGLNTSTGVQDQ 796
QY 913 TCTVENQPT-----NKPADSLPEAPNEKPV 938
Db 797 TGTQVKPSSOGGRKASQFLPSTGEKKSII 826

RESULT 14

US-09-134-001C-5080

; Sequence 5080, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lyon Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134.001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5080

; LENGTH: 3696

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5080

Query Match

Best Local Similarity 18.8%; Pred. No. 0.00033;

Matches 204; Conservative 163; Mismatches 407; Indels 309; Gaps 49;

QY 13 KDNRRVYVDGSQSS--QKSENLTDP-----QVSQEGIQAEQIVIKITDQGYVTSHG 65
Db 1911 NSDTER--EVNGAKTNGLEKINNIOPTQTNAKQKINDKAQELQIINPTDPTAT---- 1964
QY 66 HYHYNGKVPYDALFSBELLMKDPNQLKDADIVNEVKGYYIIVKDGKYYVYLKDAHAD 125
Db 1965 -----EE-----EKQEATNRVYNAQLAQAIQ-----NIN 1987
QY 126 NVRTKDEINROKQEHVKDNKVNNAVARSQGRYTTNDGVYFNPADIETGNAYIVPH 185
Db 1988 NAHSTQEVNESKTSNIATIKRSVQPNVKKPTAINSLTQEAN--NOKTLIGNDGN----- 2040
QY 186 GGHVHYIPKSDLSASELAHAKHLAGKNMOPSQLSYSTASDNNQTSV-AKGST-----S 239
Db 2041 -----TDDEKAAKQLVTKLNEQIQIKIHST-QDNQVDNVKQAQATIAIKLIN 2087
QY 240 KPANKSENLSLKLKELYDPSAORYSESGLVFPDPAKIIISRTNGVAIP-----HGDIHYF 295
Db 2088 ANAHKRODAINILTNLAESKKSIDIRANQDATTETKNTAICSIDDTLAQARNNINGANTNA 2147
QY 296 IPYSKLSALEKTIARMYPISGTGSTVSTNAKPNVSVSLG-----SLSSNPSSLTTSKE-- 349
Db 2148 LVLENEDGGKQKQIRIV-----LSTQTKTQAKA-DIAQAICQQRSTIDQONATTEKQEA 2202
QY 350 ---LSSASDGY-----IFNPKDIVEETATAYIVRHGDHPHYI-PKSNQI----- 389
Db 2203 LERLUNQETNGVNDRIQAAALANQNVTDKNNILET-----IRNVEPIVIVKPKANEIRKK 2257
QY 390 --GOPTLPNNSLATPSPSLPINTGTSHE-KHE---EDGYGFDANRIIAEDSGFVMSHGD 443
Db 2258 AAEOTTLINQNDATLEEKQIALKLEEVKNEALNQVSOAHNSNDVKIAENNGIAKISEV 2317

Search completed: May 13, 2003, 13:58:31
Job time : 36.7951 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:53:57 ; Search time 25.1868 seconds
(without alignments)
3723.142 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAYALNHRQENKDNRRVS.....IELRLPGEVIKKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	5315	100.0	1039	9	US-09-884-465A-6	Sequence 6, Appl1
2	4322	81.3	1238	9	US-09-884-465A-381	Sequence 381, App
3	4276.5	80.5	840	9	US-09-884-465A-10	Sequence 10, Appl
4	4261.5	80.2	1365	9	US-09-884-465A-382	Sequence 382, App
5	4258.5	80.1	1378	9	US-09-884-465A-378	Sequence 378, App
6	4237	79.7	1152	9	US-09-884-465A-379	Sequence 379, App
7	3384	63.7	906	9	US-09-884-465A-372	Sequence 372, App
8	3375	63.5	906	9	US-09-884-465A-332	Sequence 332, App
9	3365	63.3	906	9	US-09-884-465A-372	Sequence 372, App
10	3362	63.3	900	9	US-09-884-465A-334	Sequence 334, App
11	3345.5	62.9	901	9	US-09-884-465A-343	Sequence 343, App
12	3338	62.8	900	9	US-09-884-465A-335	Sequence 335, App
13	3329	62.6	900	9	US-09-884-465A-339	Sequence 339, App
14	3329	62.6	900	9	US-09-884-465A-341	Sequence 341, App
15	3327.5	62.6	901	9	US-09-884-465A-345	Sequence 345, App
16	3325	62.6	894	9	US-09-884-465A-337	Sequence 337, App
17	3323.5	62.5	895	9	US-09-884-465A-347	Sequence 347, App
18	3297	62.0	633	9	US-09-884-465A-350	Sequence 350, App
19	3297	62.0	633	9	US-09-884-465A-354	Sequence 354, App

20	3297	62.0	633	9	US-09-884-465A-361	Sequence 361, App
21	3297	62.0	633	9	US-09-884-465A-362	Sequence 362, App
22	3288.5	61.9	895	9	US-09-884-465A-344	Sequence 344, App
23	3288	61.9	633	9	US-09-884-465A-351	Sequence 351, App
24	3288	61.9	633	9	US-09-884-465A-352	Sequence 352, App
25	3288	61.9	633	9	US-09-884-465A-357	Sequence 357, App
26	3288	61.9	633	9	US-09-884-465A-358	Sequence 358, App
27	3288	61.9	633	9	US-09-884-465A-359	Sequence 359, App
28	3288	61.9	633	9	US-09-884-465A-333	Sequence 333, App
29	3288	61.9	900	9	US-09-884-465A-333	Sequence 333, App
30	3288	61.9	906	9	US-09-884-465A-373	Sequence 373, App
31	3287	61.8	633	9	US-09-884-465A-355	Sequence 355, App
32	3287	61.8	633	9	US-09-884-465A-363	Sequence 363, App
33	3287	61.8	633	9	US-09-884-465A-364	Sequence 364, App
34	3287	61.8	633	9	US-09-884-465A-336	Sequence 336, App
35	3281	61.7	894	9	US-09-884-465A-365	Sequence 365, App
36	3278	61.7	894	9	US-09-884-465A-342	Sequence 342, App
37	3272	61.6	894	9	US-09-884-465A-340	Sequence 340, App
38	3272	61.6	894	9	US-09-884-465A-342	Sequence 342, App
39	3270.5	61.5	895	9	US-09-884-465A-345	Sequence 345, App
40	3269	61.5	633	9	US-09-884-465A-353	Sequence 353, App
41	3269	61.5	906	9	US-09-884-465A-371	Sequence 371, App
42	3268	61.5	888	9	US-09-884-465A-338	Sequence 338, App
43	3266.5	61.5	889	9	US-09-884-465A-348	Sequence 348, App
44	3251	61.2	627	9	US-09-884-465A-366	Sequence 366, App
45	3249	61.1	627	9	US-09-884-465A-367	Sequence 367, App

ALIGNMENTS

RESULT 1
US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication NO. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 100.0%; Score 5315; DB 9; Length 1039;
Best Local Similarity 100.0%; Pred. No. 6e-302;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNHRQENKDNRRVSVDGSSQSKSENLTDPQVSKEGIOAEQIVIKITDGYV 60
Db 21 CAYALNHRQENKDNRRVSVDGSSQSKSENLTDPQVSKEGIOAEQIVIKITDGYV 80

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKDADIVNEVKGYYIKVDGKYYVLKD 120
Db 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKDADIVNEVKGYYIKVDGKYYVLKD 140

QY 121 AAHADNVRTKDEINRQKQEHVKONEKVNNAVARSQGRYTTNDGYVFNPAIDTGN 180
Db 141 AAHADNVRTKDEINRQKQEHVKONEKVNNAVARSQGRYTTNDGYVFNPAIDTGN 200

QY 181 YIVPHGHHYHYPKSDLSASELAAKAHLAGKNKQPSQLSYSTASDNNNTQSVAKGSTK 240

201	YVPRGHGYHYIPKSOLSAEAAKAHLAKNMQPSGLSYSTASDNNTQSVAKGYSK	260
241	PANKSENLOQLLXELYDSPAQRYSZSDGLVFDPAKTIISRTPNGVAIPGHDHYHETPYSK	300
261	PANKSENLOQLLXELYDSPAQRYSZSDGLVFDPAKTIISRTPNGVAIPGHDHYHETPYSK	320
301	LSALEEKIARWPVLSCTGSTVSTNAKPNEVYSSGLSSSNPSSLTTSKELSSASDGYLFN	360
321	LSALEEKIARWPVLSGTGSTVSTNAKPNEVYSSGLSSSNPSSLTTSKELSSASDGYLFN	380
361	PKDIVETATAYIVRHGDHFHYPKSNQIGOPTLPNNSLATPSPSLP INPGT SHEKHEED	420
381	PKDIVETATAYIVRHGDHFHYPKSNQIGOPTLPNNSLATPSPSLP INPGT SHEKHEED	440
421	GYGPDANKIATBDESGFVMSHGDHNNHYFFKKDLTDEOLKAAQKHLEBVKTSHNGLDLSLS	480
441	GYGPDANKIATBDESGFVMSHGDHNNHYFFKKDLTDEOLKAAQKHLEBVKTSHNGLDLSLS	500
481	HQDYYPGNAXEMKDLDKKIIEEKIAGIMKQYGVYKRESIVVYKKEKNAIIPHGDIHHADPID	540
501	HQDYYPGNAXEMKDLDKKIIEEKIAGIMKQYGVYKRESIVVYKKEKNAIIPHGDIHHADPID	560
541	EHPKVGITGHSHSNYELFKPDEEGVAKKEGKNYVTGBELTVNVNLLKNSTFNQNQFTLANGQ	600
561	EHPKVGITGHSHSNYELFKPDEEGVAKKEGKNYVTGBELTVNVNLLKNSTFNQNQFTLANGQ	620
601	KRVSPSPPELEKKLGINMLVKLIITPDGKLEKYSKGKVFEGVGNGNTANFELDOPYLPQGT	660
621	KRVSPSPPELEKKLGINMLVKLIITPDGKLEKYSKGKVFEGVGNGNTANFELDOPYLPQGT	680
661	PKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPFGAGDTYLRVNPQFAVPGKTDAL	720
681	PKYTTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPFGAGDTYLRVNPQFAVPGKTDAL	740
721	VEVDFDEFGNAYLENNYKVEIKLPTPKLNGTFTTAGNKIPTVFEMANAYLDNQSYTIVE	780
741	VEVDFDEFGNAYLENNYKVEIKLPTPKLNGTFTTAGNKIPTVFEMANAYLDNQSYTIVE	800
781	VPILEKENQTDKPSILPQFKRNKAQNSKLDKEVEEPTSEKVEKEKLSGTGNSNSTL	840
801	VPILEKENQTDKPSILPQFKRNKAQNSKLDKEVEEPTSEKVEKEKLSGTGNSNSTL	860
841	BEVPTVPDQVEKVAKFAESYGMKLENLVFNMDGTIELYLPSPGEVIKKNMADFTGEAPOQN	900
861	BEVPTVPDQVEKVAKFAESYGMKLENLVFNMDGTIELYLPSPGEVIKKNMADFTGEAPOQN	920
901	GENKPSENGKVSVTGTVEZNQPTENKPADSLPEAPNEPKPVKPEPNDSTNGMLNPEGNVGSDDPK	960
921	GENKPSENGKVSVTGTVEZNQPTENKPADSLPEAPNEPKPVKPEPNDSTNGMLNPEGNVGSDDPK	980
961	LDPALEEAPADVPQVEKLEKFTASYGLGLSDSYFNMQGTIELRLPSPGEVIKKNLSDFIA	1019
981	LDPALFEAPADVPQVEKLEKFTASYGLGLSDSYFNMQGTIELRLPSPGEVIKKNLSDFIA	1039

RESULT 2

US-09-884-465A-381
Sequence 381, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josse
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20

Db 720 EKIAGIMKQVKKRESIVYNNKNAIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPE 779
QY 561 EGVAKEGKNVYTGEEUTNVNLLKNSTNNQNTLANGOKRVSFSPPELEKKGGINML 620
Db 780 EGVAKEGKNVYTGEEUTNVNLLKNSTNNQNTLANGOKRVSFSPPELEKKGGINML 839
QY 621 VKLITPDGKLVKSVKVGEGVGNANPELDQYLPFGOTFYTTASKDYPEVSDGTFT 680
Db 840 VKLITPDGKLVKSVKVGEGVGNANPELDQYLPFGOTFYTTASKDYPEVSDGTFT 899
QY 681 VPTSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFHGNAYLENNYKV 740
Db 900 VPTSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFHGNAYLENNYKV 959
QY 741 EIKLPIKLNQGTTRAGNKIPVTFMANAYLDNQSTYIIVEPILKENTDKPSILPOK 800
Db 960 EIKLPIKLNQGTTRAGNKIPVTFMANAYLDNQSTYIIVEPILKENTDKPSILPOK 1019
QY 801 RNKAQENSLDKVEBPKTSEKVEKEKLSGTNSTNSLLEEVPVDPVOEKVAFASY 860
Db 1020 RNKAQENSLDKVEBPKTSEKVEKEKLSGTNSTNSLLEEVPVDPVOEKVAFASY 1079
QY 861 GMKLENVLFNMGTIELYLPSEGVKKNAADFTGEAPQNGENKPSNGKSTGIVENOP 920
Db 1080 GMKLENVLFNMGTIELYLPSEGVKKNAADFTGEAPQNGENKPSNGKSTGIVENOP 1139
QY 921 TENKPADSLPEAPNEKPVKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVOEKLEK 980
Db 1140 TENKPADSLPEAPNEKPVKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVOEKLEK 1199
QY 981 FTASYGLGLDSVIFNMMDGTIELRLPSGEVKKNLSDFTA 1019
Db 1200 FTASYGLGLDSVIFNMMDGTIELRLPSGEVKKNLSDFTA 1238

RESULT 3

US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

Query Match 80.5%; Score 4276.5; DB 9; Length 840;
Best Local Similarity 82.3%; Pref. No. 1.7e-241;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHRSQENKNNRVSYVDSQSSQKSENLPDQVSKQEGIQABQIVIKITDQGV 60
Db 1 CAYALNQHRSQENKNNRVSYVDSQSSQKSENLPDQVSKQEGIQABQIVIKITDQGV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKADIVNEVKGYIIVKDKYVYVILKD 120
Db 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKADIVNEVKGYIIVKDKYVYVILKD 120
QY 121 AAHADNVRTKDEINRQKOEHVKNKNSVAVARSQGYTTNDGYVFNPAIDIEDTIGNA 180

Db 121 AAHADNVRTKDEINRQKOEHVKNKNSVAVARSQGYTTNDGYVFNPAIDIEDTIGNA 180
QY 181 YIVPHGGHYHYIPKSDLSASELAARAKAHLAGKNMOPSOISYSSTASDNNTOSVAKGTSK 240
Db 181 YIVPHGGHYHYIPKSDLSASELAARAKAHLAGKNMOPSOISYSSTASDNNTOSVAKGTSK 223
QY 241 PANKSENLOSLKELYDPSAQRYSSEGLVDFPAKLIISRTPNGVAIPHGDHYHIFPYSK 300
Db 224 ----- 223
QY 301 LSALEEKIARVPIISGTSTVSTNAKNPNEVSSIGLSLSPSSLTTSKELSSASDGYIFN 360
Db 224 ----- 223
QY 361 PKDIVEETATAYIVRHGDHFHVPKSNQIOGOTPLPNNSLATPSPSLPINPGTSHKHEED 420
Db 224 -----TPSPSLPINPGTSHKHEED 243
QY 421 GYGFDAIRIIAEDSGFYMSHGDHNYFFKKDLTDEQIKAAQKHLKEEVKTSNGLDLSLS 480
Db 244 GYGFDAIRIIAEDSGFYMSHGDHNYFFKKDLTDEQIKAAQKHLKEEVKTSNGLDLSLS 303
QY 481 HEODYPGNAKEMKDLKIEEKIAGIKMKGKRESIVVNEKNAIYPHGDHHDADID 540
Db 304 HEQDYPSNAKEMKDLKIEEKIAGIKMKGKRESIVVNEKNAIYPHGDHHDADID 363
QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVTGEEELTNVNLKNSTFNNQNTLANGO 600
Db 364 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVTGEEELTNVNLKNSTFNNQNTLANGO 423
QY 601 KRVSFPPELEKKGGINMLVKLITPDGKLVKSVKVGEGVGNANFELDQYLPQGT 660
Db 424 KRVSFPPELEKKGGINMLVKLITPDGKLVKSVKVGEGVGNANFELDQYLPQGT 483
QY 661 PKYTIASKDYPEVSDGTFTVPTSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
Db 484 PKYTIASKDYPEVSDGTFTVPTSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDAL 543
QY 721 VRVDEPHGNAYLENNYKVGEEKIPIPKLNQGTTRAGNKIPVTFMANAYLDNQSTYIVE 780
Db 544 VRVDEPHGNAYLENNYKVGEEKIPIPKLNQGTTRAGNKIPVTFMANAYLDNQSTYIVE 603
QY 781 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLS 840
Db 604 VPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLS 663
QY 841 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVKKNMADFTGEAPQGN 900
Db 664 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVKKNMADFTGEAPQGN 723
QY 901 GENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNVGSDPM 960
Db 724 GENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNVGSDPM 783
QY 961 LDPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVKKNL 1014
Db 784 LDPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVKKNL 837

RESULT 4

US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044

LOCATION: (1)...(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC.FEATURE
LOCATION: (570)...(570)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC.FEATURE
LOCATION: (571)...(571)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-378

Query Match 80.1%; Score 4258.5; DB 9; Length 1378;
Best Local Similarity 63.7%; Pred. No. 3.7e-240;
Matches 879; Conservative. 37; Mismatches 61; Indels 403; Gaps 13;

QY 25 SQSSQSENL--TPDQVSQKEIQABQIVKITDQGYVTSHGHDHYHNGKVPYPDALFSE 82
DB 17 SERHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF-----IPYEQMSL 60
QY 83 E-----LMKDDNYO-----LKDA 96
DB 61 EKRIARIILYRSNHWVDSRPEQSPQSTPEPSPLOPAPNPQAPNSPIDEKLKVEA 120
QY 97 DIVNEVKGYYITKVG-KYYVYLKD-----AAHADNVRTKDEI----- 133
DB 121 --VRKVGDDGVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLGHKLGAKKTDLPSSDR 178
QY 134 -----NRQKQ-----EHVKD--NEKV----- 147
DB 179 EYFNKAYDLLARIHODLNDKGRQVDFEVLNLLERLKDVSDDKVKLVDDILAFAPIRH 238
QY 148 -----NSNAVARSQRYTNDGYFNPADIIEDTGNAXIVPHGGHYHIP 193
DB 239 PERLQKPNQAITYTDDIEQVAKLAGYTTEDGYIPDPDITSDGEDAVYTPHMTSHWIK 298
QY 194 KSDLSASELAARAKAHLAG----- 212
DB 299 KDSLEAERAAQAAYAKAKGLTTPSTDHDSQNTPEAKGAEAIYNRVKAAKVPLDRMPYN 358
QY 213 ----- 212
DB 359 LQYTVKNGSLIIPHYDHYHNKFEWDEGLYEPKGYSLLEDLLATVYVYVEHNERPH 418
QY 213 ----- 212
DB 419 SDNGFGNASDHVRKNKADQSKPDEKHEDEVSEPTHEPSEKENHAGINPSADNLKPS 478
QY 213 ----- 212
DB 479 TDTEETEBAEDTTDEALPQVENSVINAKTADAEALLEKVTDPDSIRONAMETLTGLKS 538
QY 213 -----NMQPSQLSYSTASDNNTQSVAKGSTS 239
DB 539 LLLGTKDNNTISAEDVSLIALLKESQAPAIQXNNQPSQLSYSTASDNNTQSVAKGSTS 598
QY 240 KPANKSENQLSKELYDPSAQRYSESGLVDFDPAKIISRTPNGVAIPHGCHYHFIPYS 299
DB 599 KPANKSENQLSKELYDPSAQRYSESGLVDFDPAKIISRTPNGVAIPHGCHYHFIPYS 658
QY 300 KLSALEEXIARMPISGTVSTINAKPNEVYSSLSGSSNPSLTTSKELSSASDGVI 359
DB 659 KLSALEEXIARMPISGTVSTINAKPNEVYSSLSGSSNPSLTTSKELSSASDGVI 718
QY 360 NPKDIVEETATAYIVRHGDHYPHYPKSNQIQOPLPNNSLATPSPSLPINCTSHERKEE 419
DB 719 NPKDIVEETATAYIVRHGDHYPHYPKSNQIQOPLPNNSLATPSPSLPINCTSHERKEE 778
QY 420 DQYGFANDRIIAEDSGFVMSHGDNHYFFKKDLTTEQIKAAQKHLEVKTSNGLDLSLS 479
DB 779 DQYGFANDRIIAEDSGFVMSHGDNHYFFKKDLTTEQIKAAQKHLEVKTSNGLDLSLS 838
QY 480 SHEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVWKNENAIYPHGDHHDADI 539
DB 839 SHEQDYPGNAKEMKDLKKIEKTAGIMQYGVKRESIVVWKNENAIYPHGDHHDADI 898

QY 540 DEHKPVGICHSHSNVLEFKPEGVAKKEGKNVYTGEEELTNVNLKSTFNQNFLLANG 599
DB 899 DEHKPVGICHSHSNVLEFKPEGVAKKEGKNVYTGEEELTNVNLKSTFNQNFLLANG 958
QY 600 QKRVSFSFPPELEKKGILNMLVKLITPDGKVLKSVKGVGEGVGNIANFELDQYLPQG 659
DB 959 QKRVSFSFPPELEKKGILNMLVKLITPDGKVLKSVKGVGEGVGNIANFELDQYLPQG 1018
QY 660 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLVRNPOFAVPKGTDA 719
DB 1019 TFKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLVRNPOFAVPKGTDA 1078
QY 720 LVRVFEDEFHGNAYLENNYKVGIEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIV 779
DB 1079 LVRVFEDEFHGNAYLENNYKVGIEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIV 1138
QY 780 EYVPLEKENQTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKELSTGNSNST 839
DB 1139 EYVPLEKENQTDKPSILPQFKRNKAQENSKLDEKVEPKTSEKVEKELSTGNSNST 1198
QY 840 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNMADFTGEAPQG 899
DB 1199 LEEVPTDVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNMADFTGEAPQG 1258
QY 900 NGEKKPSENGKVSCTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSGDP 959
DB 1259 NGEKKPSENGKVSCTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSGDP 1318
QY 960 MLDPALEAPADVPQVEKLEKFTASYGLGDSVFNMDGTIELRLPSGEVTKKLSDFIA 1019
DB 1319 MLDPALEAPADVPQVEKLEKFTASYGLGDSVFNMDGTIELRLPSGEVTKKLSDFIA 1378

RESULT 6

US-09-884-465A-379
Sequence 379, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: Patent in version 3.1
SEQ ID NO 379
LENGTH: 1152
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC.FEATURE
LOCATION: (1)...(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC.FEATURE
LOCATION: (344)...(344)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC.FEATURE
LOCATION: (345)...(345)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-379

Query Match 79.7%; Score 4237; DB 9; Length 1152;
Best Local Similarity 77.0%; Pred. No. 5.1e-239;
Matches 857; Conservative 36; Mismatches 92; Indels 128; Gaps 13;

QY 8 HRSQENKNNRVSYYVDSQSSQKSENLTDPQVSQKEGQAE----- 48
Db 67 HSHWIKKDSLSAEARAAQAYAKAGLTPSTDHQDSGNTAKGAEAIYNRVKAACKVPL 126
QY 49 -----QIVIKITDQGYVTHSHCDHYHYNGVPYDALFSELLAKKDPNYQLKDAIVNE 101
Db 127 DRPMYNLOQTVVEKNGSLIPHYDHYH--NIRPEW---FDEGLYAPKGYSLD----- 175
QY 102 VKGGYIKVDGKYIV-YLKDAHAADN--VRTKDEINRQK--QEHVKONEKVNKNVAVARS 156
Db 176 -----LLATVKYVVEHPNERPHSDNGFCNASDHRVKNKADQDSKPDKEHDEV-----S 225
QY 157 QGRYTND-----GYVFNPA-----DIIDTGNAYIVPHGHHYHYIPKSDL 197
Db 226 EPTHPSDEKENHAGLNPSADNLKPDSTETEEDTEAEIPIQ-----VENSVI 279
QY 198 SA-----SELAAAK----- 206
Db 280 NAKIADAELAEKVTDPSTFQNAEMETLTGLKSLLLGTGKDNNTISAEDVSLALLKESQP 339
QY 207 AHLAGNMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLOSLKELYDPSAQRYSE 266
Db 340 APIQXXNMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLOSLKELYDPSAQRYSE 399
QY 267 SDGLVDPKALISRTPNGVAIPHGDHVFHPIYPSKLSALEEKTARMVPISGTGSTVSTNAK 326
Db 400 SDGLVDPKALISRTPNGVAIPHGDHVFHPIYPSKLSALEEKTARMVPISGTGSTVSTNAK 459
QY 327 PNEVSSGLSLSNPSSLTSSKLSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKS 386
Db 460 PNEVSSGLSLSNPSSLTSSKLSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKS 519
QY 387 NQIQGQPTLPNNSLATPSPLPINFPGTSHKHEHDGFGFDANRIIAEDSGFVMSHGDH 446
Db 520 NQIQGQPTLPNNSLATPSPLPINFPGTSHKHEHDGFGFDANRIIAEDSGFVMSHGDH 579
QY 447 YFFKDKLTEEQIKAAQKHLBEVTSNGLDLSHSHQDQPGNAKEMKDKDKTEEKIAGI 506
Db 580 YFFKDKLTEEQIKAAQKHLBEVTSNGLDLSHSHQDQPGNAKEMKDKDKTEEKIAGI 639
QY 507 MKQYGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 566
Db 640 MKQYGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 699
QY 567 EGNKVTYGEELTNVNNLLKNSTFNNQNTLANGQKRVSYSPFPELEKKGIGINMLVKLITP 626
Db 700 EGNKVTYGEELTNVNNLLKNSTFNNQNTLANGQKRVSYSPFPELEKKGIGINMLVKLITP 759
QY 627 DGKVLKVSQKVGEGVGNANTANFELQYPLPGQTFKYTTASKDYPEVSDGTFTVPTSIA 686
Db 760 DGKVLKVSQKVGEGVGNANTANFELQYPLPGQTFKYTTASKDYPEVSDGTFTVPTSIA 819
QY 687 YKMASOTIFYPEHAGDYLAVNPQFVAPKGTDALRVDFDHGNAYLNNYKVEIKLPI 746
Db 820 YKMASOTIFYPEHAGDYLAVNPQFVAPKGTDALRVDFDHGNAYLNNYKVEIKLPI 879
QY 747 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENOTDKPSLTPQKRNKAQE 806
Db 880 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENOTDKPSLTPQKRNKAQE 939
QY 807 NSKLDEKVEPKTSEKYEKELKSGTGNSTNSLTLEEVPTVDPQEVKAKFAESYGNKLEN 866
Db 940 NSKLDEKVEPKTSEKYEKELKSGTGNSTNSLTLEEVPTVDPQEVKAKFAESYGNKLEN 999
QY 867 VLFNMDGTIELYLPSSGVIKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPA 926
Db 1000 VLFNMDGTIELYLPSSGVIKKNMADFTGEAPQNGENKPSKNGKSTGTVENOPTENKPA 1059
QY 927 DLSLEAPNEKPVKPNSTDNGLNPEGNVGSQDMLDPALAEAPAVDPQVKLEKFTASYG 986
Db 1060 DLSLEAPNEKPVKPNSTDNGLNPEGNVGSQDMLDPALAEAPAVDPQVKLEKFTASYG 1119

QY 987 LGDSVIFNMDGTIELRLPSGEVVIKKNLSDFIA 1019
Db 1120 LGDSVIFNMDGTIELRLPSGEVVIKKNLSDFIA 1152
RESULT 7
US-09-884-465A-370
; Sequence 370, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josée
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 370
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-370

Query Match 63.7%; Score 3384; DB 9; Length 906;
Best Local Similarity 72.7%; Pred. No. 2,1e-189;
Matches 691; Conservative 40; Mismatches 91; Indels 128; Gaps 12;

QY 148 NSNAVARSQRYITNDGYVFNPADIIETDGNAYIVPHGHHYHYIPKSDLSASLAQA 207
Db 7 DDEIQVAKLAGTYTDEGYIF-----DTS-----WIKKDSLSEARAAQA 47
QY 208 HLAGNMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLOSLKELYDPSAQRYSES 267
Db 48 YAKERGLIP-----PSTDHQDSGNTAKGA-----EAIYNRVKAACKVPL 87
QY 268 DGLVDPKALISRTPNGVAIPHGDHVFHPIYPS-----KLSALEEKTARM--VPI 314
Db 88 DRMPYNLOYTVEVKNGLSIIIPHYDHYHNKFEWDEGLYAPKGYSLDGLATKYIYVEP 147
QY 315 SGTGSTVSTN-----AKPNEVY-----SSLSGLSSN---PSSITTS 347
Db 148 RNASDHVRKNKADQDSKPDCKEHEVSEPTHPESDEKENHAGLNPSADNLKPDSTTEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVETA 369
Db 208 TEEEAEDTTDEAIEPTPSIRQNAEMETLTGLKSSLLGTGKDNNTISAEDVSLALLKESQ 267
QY 370 TAYIVRHGDHFIYIPKSNOIGOPTLPNNSLATPSPLPINFPGTSHKHEEDGYGFDANRI 429
Db 268 PAPI-----QGQIQGQPTLPNNSLATPSPLPINFPGTSHKHEEDGYGFDANRI 316
QY 430 IAEDESGFVMSHGDHNNHFFKDKLTEEQIKAAQKHLBEVTSNGLDLSHSHQDQPGNA 489
Db 317 IAEDESGFVMSHGDHNNHFFKDKLTEEQIKAAQKHLBEVTSNGLDLSHSHQDQPGNA 376
QY 490 KEMKDKDKTEEKIAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGH 549
Db 377 KEMKDKDKTEEKIAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHDADPIDEHKPVGIGH 436
QY 550 SHSNYELFKPEEGVAKKEGNKVYTGEEELTNVNNLLKNSTFNNQNTLANGQKRVSYSPFP 609
Db 437 SHSNYELFKPEEGVAKKEGNKVYTGEEELTNVNNLLKNSTFNNQNTLANGQKRVSYSPFP 496
QY 610 ELEKKGIGINMLVKLITPDGKVLKVSQKVGEGVGNANTANFELQYPLPGQTFKYTTASKD 669


```
Db 497 ELEKLGINMLVLIIPDGKVLKESGKVPGEVGNIANFELDPQYLPQGTFYITASKD 556
QY 670 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFPHG 729
Db 557 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFPHG 616
QY 730 NAYLENNYKVGELKIPKLNQCTTGTAGNKIPVTFMANAYLDNQSTYIVEVPILEKNO 789
Db 617 NAYLENNYKVGELKIPKLNQCTTGTAGNKIPVTFMANAYLDNQSTYIVEVPILEKNO 676
QY 790 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSLTLEEVPTDVP 849
Db 677 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSLTLEEVPTDVP 736
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNSDFTA 1019
Db 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNSDFTA 796
QY 910 KVSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVSDPMLDPALEAP 969
Db 797 KVSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVSDPMLDPALEAP 856
QY 970 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVKKNSDFTA 1019
Db 857 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVKKNSDFTA 906

RESULT 8
US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Jossee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332

Query Match 63.5%; Score 3375; DB 9; Length 906;
Best Local Similarity 72.6%; Pred No. 7e-189;
Matches 690; Conservative 40; Mismatches 92; Indels 128; Gaps 12;

QY 148 NGNAVARSQGYTTNDGVYFNPADIEDTGNAYIVPHGSHYIIPKSDLSASELAARA 207
Db 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEARAQA 47
QY 208 HLAGKNMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLOSLLKELYDSPAQYSES 267
Db 48 YAKEKGLTP-----PSTDHDSGNTFEAKG-----EATYNNRKAACKVPL 87
QY 268 DGLVDFPAKIIIRTPNGVAIPHCDHYHTIPYS-----KLSALEKIAMR---VPI 314
Db 88 DRMPYNLQYTVKNGSLIIPSYDHYHNLKPFNDFEGLYEPKGYSLDGLATVAKYIYEP 147
QY 315 SGTGTVSTN-----AKNEVV-----SSLSLSSN---PSLTTS 347
```

```
Db 148 RNASDHVRKANKADQDSKDEKDEKHEDEVSEPTHEPSEKENHAGLNPSADNLYKPSTDEE 207
QY 348 KE-----LSSASDGVIENPKD-----TWETA 369
Db 208 TEEEAEDTDEAEIIPGTPSIRONAMETITGLKSSLLGLTKDNNITISAEVDSLLALKESQ 267
QY 370 TAYIVRHGDHFIYIPKSNQIQOFTLPNNSLATPSPSLPINPGTSHKEHEEDGYFDANRI 429
Db 268 PAPI-----QGPQIGQFTLPNNSLATPSPSLPINPGTSHKEHEEDGYFDANRI 316
QY 430 IAEDESGFVMSHGDHNHYFFKDLTEEOIKAAQKLEBEVKTSHNGLDLSLSHEODYPNA 489
Db 317 IAEDESGFVMSHGDHNHYFFKDLTEEOIKAAQKLEBEVKTSHNGLDLSLSHEODYPNA 376
QY 490 KEMKOLDKIEBKIIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPIDIEHKPVGIGH 549
Db 377 KEMKOLDKIEBKIIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPIDIEHKPVGIGH 436
QY 550 SHSNYELKPEEGVAKKGNKVITGEELTNVNLKNTFNQNTFLANGOKRVYSFSEPP 609
Db 437 SHSNYELKPEEGVAKKGNKVITGEELTNVNLKNTFNQNTFLANGOKRVYSFSEPP 496
QY 610 ELEKLGINMLVLIIPDGKVLKESGKVPGEVGNIANFELDPQYLPQGTFYITASKD 669
Db 497 ELEKLGINMLVLIIPDGKVLKESGKVPGEVGNIANFELDPQYLPQGTFYITASKD 556
QY 670 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFPHG 729
Db 557 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFPHG 616
QY 730 NAYLENNYKVGELKIPKLNQCTTGTAGNKIPVTFMANAYLDNQSTYIVEVPILEKNO 789
Db 617 NAYLENNYKVGELKIPKLNQCTTGTAGNKIPVTFMANAYLDNQSTYIVEVPILEKNO 676
QY 790 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSLTLEEVPTDVP 849
Db 677 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKELSETGNSTNSLTLEEVPTDVP 736
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNSDFTA 1019
Db 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNSDFTA 796
QY 910 KVSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVSDPMLDPALEAP 969
Db 797 KVSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVSDPMLDPALEAP 856
QY 970 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVKKNSDFTA 1019
Db 857 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVKKNSDFTA 906

RESULT 9
US-09-884-465A-372
; Sequence 372, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Jossee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 372
; LENGTH: 906
; TYPE: PRT
```

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match 63.3%; Score 3365; DB 9; Length 906;

Best Local Similarity 72.5%; Pred. No. 2.7e-188;

Matches 689; Conservative 40; Mismatches 93; Indels 128; Gaps 12;

QY 148 NSNVAVARSGRYTTNDGYVFNADIEDTGNAYIVPHGGHYHYIPKSDLSASELAATAKA 207
DB 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEERAAQA 47
QY 208 HLAGKNQPSOLSSTASDNNTQSVAKGSTKPAKSENKSLKELVDSPSAQRYSDS 267
DB 48 YAKEGLTP-----PSTDHDSGNTAKGA-----EAIYNRVKAAKKVP 87
QY 268 DGLVDFPAKIISRTPNGVAIPHGGHYHYIPYS-----KLSALEEKIARM--VPI 314
DB 88 DEMPYNLOYTVEVKGSLIIPHYDHYHNKFEWDEGLYEAPKGYSLDILLATVYVYVEP 147
QY 315 SGTGSTVSTN-----AKPNEV-----SSLGSLSSN-----PSLITS 347
DB 148 RNASHVHRKKNKADQSKPDEDKEHVESEPTHEPSEDEKENHAGLNPSADNLYKPTDTEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVEETA 369
DB 208 TEEEAEDTTDEAEIPCTPSIRONAMETLGLKSSLLGLTKDNNTISAEDVSLALLKESQ 267
QY 370 TAYIVRGDHPHYIPKSNQIGQPTLPNNSLATPSLIPNPGTSHKHEEDGYGFDANRI 429
DB 268 PAPI-----QGPQIGQPTLPNNSLATPSLIPNPGTSHKHEEDGYGFDANRI 316
QY 430 IAEDESFGVMGSHGHNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNA 489
DB 317 IAEDESFGVMGSSGDSNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNA 376
QY 490 KEMKDLKKIEEKTAGTKMOYGVKRESIVVNEKNAIIPVPHGHHADPIDEHKPVGICH 549
DB 377 KEMKDLKKIEEKTAGTKMOYGVKRESIVVNEKNAIIPVPHGHHADPIDEHKPVGICH 436
QY 550 SHSNYELFKPEGVAKREGNKVYTGESLTNVNLLKNSTFNNGNFTLANGQKRVSFSEPP 609
DB 437 SHSNYELFKPEGVAKREGNKVYTGESLTNVNLLKNSTFNNGNFTLANGQKRVSFSEPP 496
QY 610 ELEKKGINMLVKLITPDGKLEKVGKVGEGVGNANFELDQYLPQGTQKTYTASKD 659
DB 497 ELEKKGINMLVKLITPDGKLEKVGKVGEGVGNANFELDQYLPQGTQKTYTASKD 556
QY 670 YPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 729
DB 557 YPEVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 616
QY 730 NAYLENNYKVEIKLPTKLNQGTTRTAGNKIPVTWNAAYLDNQSYIIVEVPILEKNO 789
DB 617 NAYLENNYKVEIKLPTKLNQGTTRTAGNKIPVTWNAAYLDNQSYIIVEVPILEKNO 676
QY 790 TDKPSILPQFRRKNAQENSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVPTDVP 849
DB 677 TDKPSILPQFRRKNAQENSKLDEKVEEPKTSKVEKEKLSSETGNSNSTLEEVPTDVP 736
QY 850 QEKVAKFAESYGMKLENVFNMDGTIELYLPSCGVIKKNMADTGEAPONGENKPSNG 909
DB 737 QEKVAKFAESYGMKLENVFNMDGTIELYLPSCGVIKKNMADTGEAPONGENKPSNG 796
QY 910 KVSSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPNCGNVSQDPMPLDPALEAP 969
DB 797 KVSSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPNCGNVSQDPMPLDPALEAP 856
QY 970 AVDPVOEKLEKFTASYGLGLDVSIFNMDGTIELRLPSGEVKKNLSDFTA 1019
DB 857 AVDPVOEKLEKFTASYGLGLDVSIFNMDGTIELRLPSGEVKKNLSDFTA 906

RESULT 10

US-09-884-465A-334
Sequence 334, Application US/09884465A
Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Quellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 334
LENGTH: 900

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

US-09-884-465A-334

Query Match 63.3%; Score 3362; DB 9; Length 900;

Best Local Similarity 74.7%; Pred. No. 4e-188;

Matches 699; Conservative 35; Mismatches 105; Indels 96; Gaps 15;

QY 148 NSNVAVARSGRYTTNDGYVFNADIEDTGNAYIVPHGGHYHYIPKSDLSASELAATAKA 207
DB 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEERAAQA 47
QY 208 HLAGKNQPSOLSSTASDNNTQSVAKGSTS-----KPANK-----SENLO----- 249
DB 48 YAKEGLTP-----PSTDHDSGNTAKGAELIYNVKAARKVPLDRMPYNLOIYVEYKN 102
QY 250 -SLI-----KELYDSPAQRYSESGLV-----PDPAKIISRTPNGVA-----IPH 289
DB 103 GSLLIPIPIKFEWDEGLYEAPKG--YSLEDLLATVYVPEPRNASCHVRKKNKADQSKPD 160
QY 290 GDVHVH--FTPYSKLSALEEKIARMVPISTGTSTVSTNAKPNEV-----VSSLGSLSSNPS 342
DB 161 EDKHDVSEPTHEPSEDEKENHAGLNPSADNLYKPTDTEETEEAEADTDEAEIPGTPS 220
QY 343 SLTTSKE--LSSASDGYIFNPKD-----IVEETATAYIVRGDHPHYIPKSNQ 388
DB 221 IRQNAETLTGLKSSLLGLTKDNNTISAEDVSLALLKESQAPI-----QGPQ 269
QY 389 IGQPTLPNNSLATPSLIPNPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGHNHYF 448
DB 270 IGQPTLPNNSLATPSLIPNPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGHNHYF 329
QY 449 FKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKTAGIMK 508
DB 330 FKDLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKTAGIMK 389
QY 509 QYGVKRESIVVNEKNAIIPVPHGHHADPIDEHKPVGICHSHSNYELFKPEGVAKREG 568
DB 390 QYGVKRESIVVNEKNAIIPVPHGHHADPIDEHKPVGICHSHSNYELFKPEGVAKREG 449
QY 569 NKVYTGELTNVNLKSNSTFNNGNFTLANGQKRVSFSPPELEKKGINMLVKLITPDG 628
DB 450 NKVYTGELTNVNLKSNSTFNNGNFTLANGQKRVSFSPPELEKKGINMLVKLITPDG 509
QY 629 KVLKVSCKVFGEGVGNANFELDQYLPQGTQKTYTASKDYPEVSYDGTFTVPTSLAYK 688
DB 510 KVLKVSCKVFGEGVGNANFELDQYLPQGTQKTYTASKDYPEVSYDGTFTVPTSLAYK 569
QY 689 MASOTIFYPPHAGDTYLRVNPQFAPVPGTDLVRVDFEFGNAYLENNYKVEIKLPTPK 748

Db	570	MASOTIFYPHAGDTYLRVNPOAFVKGTDALVRVDFEFGHAYLENNYKVGEIKLPIPK	629
QY	749	LNQGTTRTAGNKLPVTFMANAYLDNOSTYIVFVPILEKENQTDKPSILPQFVRNKAQNS	808
Db	630	LNQGTTRTAGNKLPVTFMANAYLDNOSTYIVFVPILEKENQTDKPSILPQFVRNKAQNS	689
QY	809	KLDEKVEEPTSEKVKPEKLSLSETSGNSTSLTEEVPVDPVQEKVAKFAESYGMKLENLV	868
Db	690	KLDEKVEEPTSEKVKPEKLSLSETSGNSTSLTEEVPVDPVQEKVAKFAESYGMKLENLV	749
QY	869	FNMDGTIELYLPSEVITKKNMADFGEAPQNGENKPKSPENGKVGSTGTVENQTEKNPKADS	928
Db	750	FNMDGTIELYLPSEVITKKNMADFGEAPQNGENKPKSPENGKVGSTGTVENQTEKNPKADS	809
QY	929	LPEAPNEKPKVPENSTNDGMLNPENGVNDDPMLDPALEAPADVPOEKLKFTFASYGLG	988
Db	810	LPEAPNEKPKVPENSTNDGMLNPENGVNDDPMLDPALEAPADVPOEKLKFTFASYGLG	869
QY	989	LDSVIFNMDDGTIELRPPSGEVIKKNLSDFIA	1019
Db	870	LDSVIFNMDDGTIELRPPSGEVIKKNLSDFIA	900

```

RESULT 11
US-09-884-465A-343
; Sequence 343, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 343
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

```

208	TEEAEADTTDEAEITPGTPTSTQNAMETLTGLKSSLTGLTGKDNNTTISAEVDSLLALKESQ	26
380	PHYTPKSNQIQOPTLPNNSTATSPSLP INPGTSHKEHEDGYGFQDANRIIAEDSGFVM	439
268	PAPIQSGPQIQOPTLPNNSTATSPSLP INPGTSHKEHEDGYGFQDANRIIAEDSGFVM	327
440	SHGDHNHYFFKKDLTBEQIKAAQKHLEEVKTSINGLDSLSHEDQYPGNAKEMKDLDKKI	499
328	S-----YFFKKDLTBEQIKAAQKHLEEVKTSINGLDSLSHEDQYPGNAKEMKDLDKKI	381
500	BEKIAGIMKQYGVKRESIVYNVEKNALIIYPHGDSHHADPIDEHKPVGIGHSHSNYELFKP	559
382	EKIAGIMKQYGVKRESIVYNVEKNALIIYPHGDSHHADPIDEHKPVGIGHSHSNYELFKP	441
560	BEGVAKREGNKVYTGEBLTINVNLLKNSTFNQNFETLANGQKRVSFSPPELEKKLGJNM	619
442	BEGVAKREGNKVYTGEBLTINVNLLKNSTFNQNFETLANGQKRVSFSPPELEKKLGJNM	501
620	LVKLITPDGKVLKRVSKGVFGEVGVGNIANFELDQYLPQOTERYTITASKDYPEVSDVGTG	679
502	LVKLITPDGKVLKRVSKGVFGEVGVGNIANFELDQYLPQOTERYTITASKDYPEVSDVGTG	561
680	TVPTSLAYKNASQITFYFFHAGDTYLRVNPQFQVPGKTDALYRVFDEFHGNAYLENNKYV	739
562	TVPTSLAYKNASQITFYFFHAGDTYLRVNPQFQVPGKTDALYRVFDEFHGNAYLENNKYV	621
740	GEIKLPIPKLNQGTTRTAGNKKIPVTFMANAYLDNQSTYIVEVPILEKNOTDKPSILPQF	799
622	GEIKLPIPKLNQGTTRTAGNKKIPVTFMANAYLDNQSTYIVEVPILEKNOTDKPSILPQF	681
800	KRNAQNSKLDEKVBEPKTSSEKVEKSLSETGNSNSTLEEVPTVPQEVKAKPAES	859
682	KRNAQNSKLDEKVEEPEKTSSEKVEKSLSETGNSNSTLEEVPTVPQEVKAKPAES	741
860	YGMKLENVLENMDGTIELYLPSEVGIKKNAADFTGAPQGNKPKSNGKSVSTGTIVNQ	919
742	YGMKLENVLENMDGTIELYLPSEVGIKKNAADFTGAPQGNKPKSNGKSVSTGTIVNQ	801
920	PTENKPADSLPEAPNEKVPKPENSTDNGMLNPGENVGSDPMLDPALEAPDVPQEKLE	979
802	PTENKPADSLPEAPNEKVPKPENSTDNGMLNPGENVGSDPMLDPALEAPDVPQEKLE	861
980	KFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA	1019
862	KFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA	901

```

RESULT 12
US-09-884-465A-335
: Sequence 335, Application US/09884465A
: Publication No. US20030077293A1
: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 335
: LENGTH: 900
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
US-09-884-465A-335

```

Publication No. US20030077293A1									
GENERAL INFORMATION:									
APPLICANT: Shire Biochem, Inc.									
APPLICANT: Hamel, Josee									
APPLICANT: Brodeur, Bernard									
APPLICANT: Martin, Denis									
APPLICANT: Chellett, Catherine									
TITLE OF INVENTION: Streptococcus Antigens									
FILE REFERENCE: 055190-0044									
CURRENT APPLICATION NUMBER: US/09/884,465A									
CURRENT FILING DATE: 2001-06-20									
PRIOR FILING DATE: 2000-06-20									
NUMBER OF SEQ ID NOS: 384									
SOFTWARE: Patentin version 3.1									
SEQ ID NO 339									
LENGTH: 900									
TYPE: PRT									
ORGANISM: Artificial Sequence									
FEATURE:									
OTHER INFORMATION: Unknown Organism									
US-09-884-465A-339									
Query Match 62.6%; Score 3329; DB 9; Length 900;									
Best Local Similarity 72.1%; Pred. No. 3.3e-186;									
Matches 686; Conservative 40; Mismatches 91; Indels 134; Gaps 13;									
QY	148	NSNVAARSQGRYTTNDGYVFNPAIDTGNAYIVPHGCHYHYIPKSDLSASELAACA	207						
DB	7	DDEIQVAKLAGYTTEDGYIF-----DTS-----WIKKDSLSEAEAAQA	47						
QY	208	HLACKNMOPQSLSVSSASDNTQSVAKGSTSKPANKSENLOSLLLKELYDPSAQRYSES	267						
DB	48	YAKKGLTP-----PSTDHDSGNTKARGA-----EAIYNRVKAARKVPL	87						
QY	268	DGLVDPKAKIISRTNGVAIPHGDHYHFIPIYS-----KLSALEEKIARM---	314						
DB	88	DRMPYNLQYIVVEKNGSLIIPSYDSYHNKFEWDEGLYEAPKGYSLDGLATVYIYVEP	147						
QY	315	SGTGSTVSTN-----AKPNEV-----SSLSGSSN-----PSSLTTS	347						
DB	148	RNASDHVRKNRADOQSKPDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSDTTEE	207						
QY	348	KE-----LSSASDGYIFNPKD-----IVEETA	369						
DB	208	TEEAEDTTDEAEIPGTPSIQNAWETLGLKSSLLTCKDNNTISAEDVSLALLKESQ	267						
QY	370	TAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDANRI	429						
DB	268	PAPI-----QGPOIGOPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDANRI	316						
QY	430	IAEDSGFVMSHGDNHVFYFKDLTEQIKAAQKHLSEVKTSHNGDLSLSHEDQYPCNA	489						
DB	317	IAEDSGFVMS-----YFFKDLTEQIKAAQKHLSEVKTSHNGDLSLSHEDQYPCNA	370						
QY	490	KEMKDLKKIEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGH	549						
DB	371	KEMKDLKKIEKTAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPIDEHKPVGIGH	430						
QY	550	SHSNVELFKPEGVAKKNGKYYTGEEELTNVNNLLKSTFNNQNTLANGOKRVSFSPFP	609						
DB	431	SHSNVELFKPEGVAKKNGKYYTGEEELTNVNNLLKSTFNNQNTLANGOKRVSFSPFP	490						
QY	610	ELEKKGINMLVKLITPDGKVLKESGVKGVGEGVGNIANFELDQPYLPQTFKYTIASKD	669						
DB	491	ELEKKGINMLVKLITPDGKVLKESGVKGVGEGVGNIANFELDQPYLPQTFKYTIASKD	550						
QY	670	YPEVSYDGTFTVPTSLAYKMASQITIFYPFHAGDTYLRVNPQFAVPKGTDALRVFDEFG	729						
DB	551	YPEVSYDGTFTVPTSLAYKMASQITIFYPFHAGDTYLRVNPQFAVPKGTDALRVFDEFG	610						
QY	730	NAYLENNYKVGGEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEPFILEXNQ	789						

```

Db 611 NAYLENNYKVGELKPIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 670
QY 790 TDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNGSTNSTLEEVPTDVPV 849
Db 671 TDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNGSTNSTLEEVPTDVPV 730
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENG 909
Db 731 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENG 790
QY 910 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGVNPGSDPMLDPALEAP 969
Db 791 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGVNPGSDPMLDPALEAP 850
QY 970 AVDPVQEKLEKFTASGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 851 AVDPVQEKLEKFTASGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 900

```

RESULT 14

```

US-09-884-465A-341
; Sequence 341, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US/09/884.465A
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 341
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-341

```

```

Query Match 62.6%; Score 3329; DB 9; Length 900;
Best Local Similarity 72.1%; Pred. No. 3.3e-186;
Matches 685; Conservative 40; Mismatches 91; Indels 134; Gaps 13;

```

```

QY 148 NSNVAVSQGRYTTNDGYVFNPAADIETGNAYIVPHGGHYHYTPKSDLSASELAARAKA 207
Db 7 DDEIQVAKLAGRYTTEDGYIF-----DTS-----WIKKDSLSEAEERAAQA 47
QY 208 HLAGKNMPSQSLSTASDNSTQSVAKGSKTPANKSENLSLKLKELYDSPSAQRYSES 267
Db 48 YAKEGLAP-----PSTDHQDSGNETAKA-----EAYNRVKAQKVP 87
QY 268 DGLVDPAKILSTPNQVAIPHGDHYHYTPYS-----KLSAEKTIARM---VPI 314
Db 88 DRMPYNLOQTVKNGSLIIPSYDHYSNIKPFEWDEGLYAPKGYSLDALLATVKYVYEP 147
QY 315 SCTGTSTVSTN-----AKPNEV-----SSLGSLSSN---PSSLTTS 347
Db 148 RNASHVRKKNKADQSKPDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSSTDEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVEETA 369
Db 208 TEEAEADTTDEAEIPCTPSIRONAMETLGLKSSLLGTCKDNNTISAEVDSILLALKESQ 267
QY 370 TAYIVRHGDHFHYPKSNQIQOPTLPNNSLATPSPSLPINDGTSHKEHEEDGYGFDANRI 429

```

```

Db 268 PAPI-----QGPOIQOPTLPNNSLATPSPSLPINDGTSHKEHEEDGYGFDANRI 316
QY 430 IAEDESQVMSGHGDNHYFFKKDLTEEQIKAAQKHLSEVKTSHNGLDSLSSHEQDYPCNA 489
Db 317 IAEDESQVMS-----YFFKKDLTEEQIKAAQKHLSEVKTSHNGLDSLSSHEQDYPCNA 370
QY 490 KEMDLDKKIEEKTAGIMKQYGVKRESIVVKNENKNAIYPHGDHHDHADPIDEHKPVGIGH 549
Db 371 KEMDLDKKIEEKTAGIMKQYGVKRESIVVKNENKNAIYPHGDHHDHADPIDEHKPVGIGH 430
QY 550 SHSNVYELFKPEEGVAKKGNKVYTGEEELTNVNLKKNSTENNQNFTILANGKRVSEFSEFP 609
Db 431 SHSNVYELFKPEEGVAKKGNKVYTGEEELTNVNLKKNSTENNQNFTILANGKRVSEFSEFP 490
QY 610 ELEKKLGINMLVKLITPDGKVLKESKGVFGEVGNANTANFELDQPYLPQGTFKTYTIASKD 669
Db 491 ELEKKLGINMLVKLITPDGKVLKESKGVFGEVGNANTANFELDQPYLPQGTFKTYTIASKD 550
QY 670 YPEVSYDGTFTVPTSLAYKMASQITFIYFFHAGDTYLYRVNPOFAVPKGTDALVRVDFEHG 729
Db 551 YPEVSYDGTFTVPTSLAYKMASQITFIYFFHAGDTYLYRVNPOFAVPKGTDALVRVDFEHG 610
QY 730 NAYLENNYKVGELKPIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 789
Db 611 NAYLENNYKVGELKPIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 670
QY 790 TDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNGSTNSTLEEVPTDVPV 849
Db 671 TDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNGSTNSTLEEVPTDVPV 730
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENG 909
Db 731 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSENG 790
QY 910 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGVNPGSDPMLDPALEAP 969
Db 791 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGVNPGSDPMLDPALEAP 850
QY 970 AVDPVQEKLEKFTASGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 851 AVDPVQEKLEKFTASGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 900

```

RESULT 15

```

US-09-884-465A-345
; Sequence 345, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 345
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

```

```

Query Match 62.6%; Score 3327.5; DB 9; Length 901;
Best Local Similarity 72.9%; Pred. No. 4.1e-186;
Matches 685; Conservative 37; Mismatches 105; Indels 113; Gaps 12;

```

